Wahrscheinlichkeitstheorie

Peter Eichelsbacher

Wintersemester 2016/17

Inhaltsverzeichnis

KAPITEL 1

σ-Algebren und Maße

Wir starten in diesem Kapitel mit der sogenannten *Maßtheorie*, der allgemeinen Theorie des Messens von Inhalten von Strecken, Flächen, Körpern und Mengen in höherdimensionalen Räumen. Mit Hilfe des CAUCHY–RIEMANNschen Integrals kann man Gebieten zwischen dem Graphen einer speziellen Klasse von Funktionen und der entsprechenden Achse einen Flächeninhalt zuweisen. Man möchte einer möglichst großen Klasse von Bereichen, insbesondere des $\mathbb{R}^d$, in einer sinnvollen Art und Weise einen Inhalt zuordnen. Gesucht ist eine Teilmenge $\mathcal{A}$ der Potenzmenge $\mathcal{P}(\mathbb{R}^d)$ und eine Abbildung $\mu : \mathcal{A} \to [0, \infty]$, so dass für jedes $A \in \mathcal{A}$ der Wert $\mu(A)$ als Inhalt interpretiert werden kann. Gewünscht ist dabei natürlich, dass den bekannten elementargeometrischen Figuren der vertraute Inhalt zugeordnet wird und dass zum Beispiel der Inhalt der Vereinigung zweier disjunkter Bereiche gleich der Summe der Inhalte der einzelnen Bereiche sei. Man wünscht sich im $\mathbb{R}^d$ auch die Unabhängigkeit des Inhalts einer Menge von der Lage im Raum. Es wird sich später herausstellen, dass man ein “d-dimensionales Volumen” nicht für alle Teilmengen, sondern nur für eine bestimme Auswahl von Teilmengen definieren kann.

Wir führen zunächst diejenigen Mengensysteme ein, auf denen wir dann Inhalte bzw. Maße erklären werden. Es wirkt plausibel, dass man sich von einem geeigneten Mengensystem wünscht, dass es abgeschlossen ist gegenüber allen abzählbaren Mengenoperationen. Dabei heißt formal ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$, Ω ist eine nichtleere Menge, abgeschlossen unter allen abzählbaren Mengenoperationen, wenn mit $A \in \mathcal{A}$ immer $A^c \in \mathcal{A}$ folgt und wenn für jede Folge $(A_n)_n$ in $\mathcal{A}$ auch $\bigcup_{n\geq 1} A_n$ zu $\mathcal{A}$ gehört. Der Begriff ist dadurch gerechtfertigt, dass aufgrund der Regeln von DE MORGAN somit auch $\bigcap_{n\geq 1} A_n$ zu $\mathcal{A}$ gehört.

Wir führen die Begriffe Inhalt, Prämaß und Maß ein und leiten Rechenregeln her, die für den weiteren Aufbau der Maßtheorie und der im Anschluß zu entwickelnden Integrationstheorie die Grundlage darstellen. Als wichtiges Beispiel führen wir das LEBESGUEsche Prämaß ein.

1.1 Definition Ein System $\mathcal{A}$ von Teilmengen einer Menge Ω heißt eine σ-*Algebra* (in Ω), wenn gilt:

(a) $\Omega \in \mathcal{A}$

(b) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

(c) für jede Folge $(A_n)_n$ von Mengen aus $\mathcal{A}$ liegt $\bigcup_{n=1}^{\infty} A_n$ in $\mathcal{A}$.

Das Paar $(\Omega, \mathcal{A})$ heißt *Messraum*.

1.2 Beispiele

(a) $\mathcal{P}(\Omega)$ ist eine σ-Algebra.

(b) Ist $\mathcal{A}$ eine σ-Algebra in Ω und $\Omega' \subset \Omega$, so ist $\Omega' \cap \mathcal{A} = \{\Omega' \cap A : A \in \mathcal{A}\}$ eine σ-Algebra in Ω' und heißt *Spur* oder Spur-σ-Algebra von $\mathcal{A}$ in Ω'.

(c) Es seien Ω und Ω' Mengen, $\mathcal{A}'$ eine σ-Algebra in Ω' und $T : \Omega \to \Omega'$ sei eine Abbildung von Ω nach Ω'. Dann ist $T^{-1}(\mathcal{A}') := \{T^{-1}(A'), A' \in \mathcal{A}'\}$ eine σ-Algebra in Ω (die durch T induzierte). Dies ist eine einfache Übung.

(d) Es sei $\mathcal{A}$ eine σ-Algebra. Dann ist $\emptyset \in \mathcal{A}$ und für jede Folge $(A_n)_n$ von Mengen aus $\mathcal{A}$ ist $\bigcap_{n=1}^{\infty} A_n$ in $\mathcal{A}$, denn: $\emptyset = \Omega^c$ und $\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c$. Weiter ist mit $A, B \in \mathcal{A}$ auch $A \setminus B = A \cap B^c \in \mathcal{A}$, denn endliche Vereinigungen und Durchschnitte von Elementen aus $\mathcal{A}$ sind in $\mathcal{A}$.

(e) Sei $\mathcal{E} \subset \mathcal{P}(\Omega)$ ein nichtleeres Mengensystem. Dann ist

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{E} \subset \mathcal{A} \\ \mathcal{A}\ \sigma-\text{Algebra}}} \mathcal{A}$$

eine σ-Algebra, die man die von $\mathcal{E}$ *erzeugte* σ-Algebra nennt. Dies ist die kleinste σ-Algebra, die $\mathcal{E}$ enthält. $\mathcal{E}$ heißt *Erzeuger* von $\sigma(\mathcal{E})$ (alles ist wohldefiniert, denn $\mathcal{P}(\Omega)$ ist immer eine σ-Algebra, die $\mathcal{E}$ umfasst). Der Beweis ist einfach.

(f) $\{A \subset \Omega, A \text{ oder } A^c \text{ ist abzählbar}\}$ ist eine σ-Algebra.

1.3 Definition Ein System $\mathcal{R}$ von Teilmengen von Ω heißt *Ring* (in Ω), wenn gilt

(a) $\emptyset \in \mathcal{R}$

(b) $A, B \in \mathcal{R} \Rightarrow A \setminus B \in \mathcal{R}$

(c) $A, B \in \mathcal{R} \Rightarrow A \cup B \in \mathcal{R}$.

Gilt auch

(d) $\Omega \in \mathcal{R}$,

so heißt $\mathcal{R}$ eine *Algebra* (in Ω).

1.4 Bemerkungen

(a) Da $A \cap B = A \setminus (A \setminus B)$, ist in einem Ring $\mathcal{R}$ mit $A, B \in \mathcal{R}$ auch $A \cap B \in \mathcal{R}$.

(b) $\mathcal{R}$ ist genau dann eine Algebra, wenn (a) und (b) aus 1.1 und (c) aus 1.3 gelten. Denn ist $\mathcal{R}$ eine Algebra, so sind (a) in 1.1 und (c) in 1.3 klar. Aus (b) in 1.3 folgt (b) in 1.1. Umgekehrt gilt $A \setminus B = A \cap B^c = (A^c \cup B)^c$ sowie $\emptyset = \Omega^c$.

(c) Es sei $\mathcal{R}$ eine Algebra und für jede disjunkte Folge $(A_n)_n$ in $\mathcal{R}$ gelte $\bigcup_{n=1}^\infty A_n \in \mathcal{R}$. Dann ist $\mathcal{R}$ eine σ-Algebra. Hierbei bedeutet disjunkte Folge, dass $A_j \cap A_k = \emptyset$ für alle $j, k \in \mathbb{N}$ mit $j \neq k$ gilt. Der *Beweis* dazu: Es sei $(B_n)_n \in \mathcal{R}^{\mathbb{N}}$. Setze $A_1 := B_1$ und $A_{j+1} := B_{j+1} \setminus \bigcup_{k=1}^j B_k, j \in \mathbb{N}$. Dann ist $(A_n)_n$ disjunkt und $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n$. Nach Voraussetzung ist $\bigcup_{n=1}^\infty A_n \in \mathcal{R}$. $\square$

1.5 Beispiele

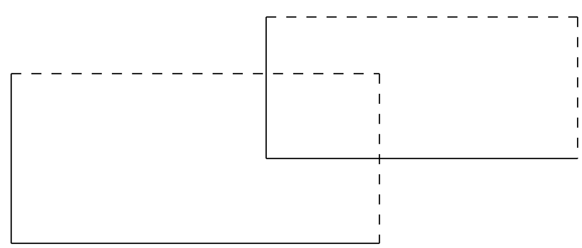
(a) σ-Algebren sind Algebren.

(b) Der kleinste in Ω existierende Ring besteht nur aus der leeren Menge.

(c) $\{A \subset \Omega, A \text{ oder } A^c \text{ endlich }\}$ ist eine Algebra, und eine σ-Algebra genau dann, wenn Ω endlich ist.

Sei nun $\Omega = \mathbb{R}^d$, $a = (a_1, \ldots, a_d)$, $b = (b_1, \ldots, b_d)$. Wir definieren $a \leq b$ bzw. $a < b$, wenn $a_i \leq b_i$ bzw. $a_i < b_i$ für alle $i = 1, \ldots, d$ gilt. Betrachte das nach rechts halboffene Intervall $[a, b) := \{x \in \mathbb{R}^d \colon a \leq x < b\}$. Dies nennt man auch einen achsenparallelen, nach rechts hin offenen Quader oder auch ein *Parallelotop*. $\mathcal{R}^d$ bezeichne die Menge aller nach rechts halboffenen Intervalle in $\mathbb{R}^d$ und $\mathcal{F}^d$ das System aller Vereinigungsmengen von je endlich vielen Mengen aus $\mathcal{R}^d$. Dieses System wird als das System der d-dimensionalen *Figuren* bezeichnet. Sicher gilt $\mathcal{R}^d \subset \mathcal{F}^d$. Es gilt weiter:

1.6 Lemma *Mit $I, J \in \mathcal{R}^d$ gilt $I \cap J \in \mathcal{R}^d$ und $J \setminus I \in \mathcal{F}^d$. Jede Figur ist endliche Vereinigung paarweise disjunkter Intervalle aus $\mathcal{R}^d$.*



Beweis: Es seien $I = [a, b)$ und $J = [a', b')$. Wir setzen $e_i := \max(a_i, a_i')$ und $f_i := \min(b_i, b_i')$ für $i = 1, \ldots, d$ sowie $e = (e_1, \ldots, e_d)$ und $f = (f_1, \ldots, f_d)$. Dann ist $I \cap J = [e, f)$ für $e \leq f$ und $I \cap J = \emptyset$ sonst, also $I \cap J \in \mathcal{R}^d$. Weiter gilt $J \setminus I = J \setminus (I \cap J)$. Daher zeigen wir den zweiten Teil der Behauptung für $I \subset J$ und $I \neq \emptyset$. Also ist ohne Einschränkung $a' \leq a < b \leq b'$. J_k sei eines der 3 disjunkten Intervalle $[a_i', a_i), [a_i, b_i)$ oder $[b_i, b_i')$, $i = 1, \dots, d$. J ist die Vereinigung der 3^d disjunkten Intervalle $J_1 \times \cdots \times J_d$, die sich bei Auswahl aller 3 Möglichkeiten für $J_1, \dots, J_d$ ergeben. Für $J_k = [a_k, b_k)$, $k = 1, \dots, d$, ergibt sich I, also ist $J \setminus I$ die disjunkte Vereinigung der übrigen $3^d - 1$ Intervalle der Form $J_1 \times \cdots \times J_d$. Es sei nun $F \in \mathcal{F}^d$. Dann existieren $I_1, \dots, I_k \in \mathcal{R}^d$ mit $F = I_1 \cup \cdots \cup I_k$. Es ist $F = I_1 \cup (I_2 \setminus I_1) \cup (I_3 \setminus I_1 \cup I_2) \cup \cdots \cup I_k \setminus (I_1 \cup \cdots \cup I_{k-1})$, wobei dies nun eine disjunkte Vereinigung ist. Es ist zu zeigen: Mengen der Form $I \setminus (J_1 \cup \cdots \cup J_m)$ sind disjunkte Vereinigungen von Elementen aus $\mathcal{R}^d$. Es gilt $I \setminus (J_1 \cup \cdots \cup J_m) = \bigcap_{i=1}^m (I \setminus J_i)$. Nun kann $I \setminus J_i$ als disjunkte Vereinigungen endlich vieler Intervalle aus $\mathcal{R}^d$ dargestellt werden, s.o.. Wir verwenden abschließend die Distributivität und die Tatsache, dass der Durchschnitt von Intervallen aus $\mathcal{R}^d$ wieder in $\mathcal{R}^d$ liegt. Somit ist alles bewiesen. □

Wir lernen nun einen für das Folgende sehr wichtigen Ring kennen.

1.7 Satz *$\mathcal{F}^d$ ist ein Ring.*

Beweis: Zu zeigen ist: sind $F, G \in \mathcal{F}^d$, so ist $F \setminus G \in \mathcal{F}^d$. Dazu sei

$$F = \bigcup_{i=1}^m I_i', \quad G = \bigcup_{j=1}^n I_j''.$$

Dann ist

$$F \setminus G = \bigcup_{i=1}^m \left(\bigcap_{j=1}^n (I_i' \setminus I_j'') \right).$$

Es ist zu zeigen: $\bigcap_{j=1}^n I_i' \setminus I_j''$ ist eine Figur. Nach 1.6 ist $I_i' \setminus I_j''$ eine Figur. Wir zeigen nun noch, dass der Durchschnitt zweier Figuren wieder eine Figur ist. Aber mit obiger Darstellung ist

$$F \cap G = \bigcup_{i=1}^m \bigcup_{j=1}^n (I_i' \cap I_j''),$$

und $I_i' \cap I_j'' \in \mathcal{R}^d$, also ist $F \cap G$ eine Figur. □

1.8 Bemerkung Ist $\mathcal{R}$ ein Ring in $\mathbb{R}^d$ mit $\mathcal{R}^d \subset \mathcal{R}$, so folgt $\mathcal{F}^d \subset \mathcal{R}$: $\mathcal{F}^d$ ist der von $\mathcal{R}^d$ erzeugte Ring. Dabei ist der *erzeugte Ring* wie in Beispiel 1.2 (e) definiert. Für jedes System $\mathcal{E}$ von Teilmengen einer Menge Ω existiert ein kleinster Ring $\varrho(\mathcal{E})$ in Ω welcher $\mathcal{E}$ enthält. Er heißt der von $\mathcal{E}$ erzeugte Ring.

Im Folgenden führen wir noch den Begriff eines DYNKIN-Systems ein:

1.9 Definition Ein DYNKIN-*System* $\mathcal{D}$ (über einer Menge Ω) ist ein System von Teilmengen von Ω, welches die folgenden Eigenschaften erfüllt:

(a) $\Omega \in \mathcal{D}$.

(b) $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$.

(c) Für jede Folge $(A_n)_n$ *paarweise disjunkter* Mengen aus $\mathcal{D}$ ist $\bigcup_{n\geq 1} A_n$ in $\mathcal{D}$.

Der Grund für die Einführung dieses Begriffs ist, dass (c) in Definition 1.9 häufig leichter nachweisbar ist als (c) in Definition 1.1. Es stellt sich die Frage: Wann ist ein DYNKIN-System eine σ-Algebra? Fortan nennen wir ein Mengensystem $\mathcal{E}$, welches mit je zwei Mengen - und damit auch je endlich vielen Mengen - deren Durchschnitt bzw. Vereinigung enthält, kurz *durchschnittsstabil* (in Zeichen $\cap$-*stabil*) bzw. *vereinigungstabil* (in Zeichen $\cup$-*stabil*).

1.10 Satz *Ist ein* DYNKIN-*System durchschnittsstabil, so ist es eine σ-Algebra.*

Beweis: $\mathcal{D}$ sei ein durchschnittsstabiles DYNKIN-System. Wir müssen zeigen, dass $\mathcal{D}$ abgeschlossen gegenüber abzählbaren Vereinigungen ist. Sei $(A_i)_i$ eine Folge in $\mathcal{D}$.

$$\begin{aligned} B_1 &:= A_1 \,, \\ B_n &:= A_n \backslash (A_1 \cup \cdots \cup A_{n-1}) \,; \quad n \geq 2 \,. \end{aligned}$$

Wir zeigen mit Induktion nach n, dass B_n und $A_1 \cup \cdots \cup A_n$ zu $\mathcal{D}$ gehören. Für $n = 1$ ist nichts zu zeigen. Sei $n \geq 2$. B_n hat die Darstellung $B_n = A_n \cap ((A_1 \cup \cdots \cup A_{n-1})^c)$. Per Induktionsvoraussetzung ist $A_1 \cup \cdots \cup A_{n-1}$ in $\mathcal{D}$, also auch das Komplement. Da $\mathcal{D}$ durchschnittsstabil ist, folgt $B_n \in \mathcal{D}$. $A_1 \cup \cdots \cup A_{n-1}$ und B_n sind disjunkt und $A_1 \cup \cdots \cup A_n = (A_1 \cup \cdots \cup A_{n-1}) \cup B_n$. Es gilt $A_1 \cup \cdots \cup A_n \in \mathcal{D}$. Die B_n sind paarweise disjunkt und $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$, also $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{D}$. □

1.11 Satz *Ist $\mathcal{C}$ ein durchschnittsstabiles Mengensystem in Ω, so gilt $d(\mathcal{C}) = \sigma(\mathcal{C})$, wobei $d(\mathcal{C})$ das kleinste von $\mathcal{C}$ erzeugte* DYNKIN-*System bezeichnet.*

Beweis: Es folgt sofort $d(\mathcal{C}) \subset \sigma(\mathcal{C})$, denn jede σ-Algebra ist auch ein DYNKIN-System. Es bleibt zu zeigen, dass $d(\mathcal{C})$ eine σ-Algebra ist. Dazu zeigen wir mit Satz 1.10, dass $d(\mathcal{C})$ durchschnittsstabil ist. Definiere

$$\mathcal{A} := \{A \subset \Omega : A \cap C \in d(\mathcal{C}) \; \forall C \in \mathcal{C}\} \,.$$

Da $\mathcal{C}$ durchschnittsstabil ist, folgt $\mathcal{C} \subset \mathcal{A}$. Wir zeigen, dass $\mathcal{A}$ die DYNKIN-

Eigenschaften hat: (a) ist klar. (b):

$$\begin{aligned} A \in \mathcal{A} &\Rightarrow A \cap C \in d(\mathcal{C})\ \forall C \in \mathcal{C} \\ &\Rightarrow A^c \cap C = (C^c \cup (A \cap C))^c \in d(\mathcal{C})\ \forall C \in \mathcal{C} \\ &\Rightarrow A^c \in \mathcal{A} \quad \text{(beachte: } C^c \text{ und } A \cap C \text{ sind disjunkt).} \end{aligned}$$

(c): $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, seien paarweise disjunkt. Wegen $A_n \cap C \in d(\mathcal{C})\ \forall C \in \mathcal{C}$ folgt $\left(\bigcup_{n\geq 1} A_n\right) \cap C \in d(\mathcal{C})\ \forall C \in \mathcal{C}$, d. h. $\bigcup A_n \in \mathcal{A}$. Also gilt $d(\mathcal{C}) \subset \mathcal{A}$. Wir definieren

$$\bar{\mathcal{A}} := \{A \subset \Omega : A \cap A' \in d(\mathcal{C}) \text{ für alle } A' \in d(\mathcal{C})\}\ .$$

Nun ist nach dem vorangegangenen Schritt $\mathcal{C} \subset \bar{\mathcal{A}}$. Man zeigt nun analog zu eben, dass $\bar{\mathcal{A}}$ ein Dynkin-System ist. Damit folgt $d(\mathcal{C}) \subset \bar{\mathcal{A}}$, also ist $d(\mathcal{C})$ durchschnittstabil, was zu zeigen war. □

Wir leiten aus dem letzten Satz ein praktisches Verfahren ab, welches man Dynkin-*System-Argument* nennt:

Gegeben sei $(\Omega, \mathcal{A})$, ein Messraum, und eine Aussage $(*)$, deren Gültigkeit für alle $A \in \mathcal{A}$ behauptet wird. Es gebe einen durchschnittstabilen Erzeuger $\mathcal{E}$ von $\mathcal{A}$ derart, dass $(*)$ für alle $A \in \mathcal{E}$ nachweisbar ist. Betrachte dann

$$\mathcal{D} := \{A \in \mathcal{A} : A \text{ genügt der Behauptung } (*)\}\ .$$

Zeige, dass $\mathcal{D}$ ein Dynkin-System bildet. Dann folgt aus $\mathcal{E} \subset \mathcal{D} \subset \mathcal{A}$ und $d(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{A}$ nach Satz 1.11 die Inklusionskette

$$\mathcal{A} = d(\mathcal{E}) \subset \mathcal{D} \subset \mathcal{A}\ ,$$

also $\mathcal{A} = \mathcal{D}$, also ist die Behauptung $(*)$ für alle $A \in \mathcal{A}$ bewiesen! Dieses Argument wird häufig verwendet, wie wir noch sehen werden.

Im folgenden seien Ω eine nichtleere Menge und $[0, \infty] := \mathbb{R}^+ \cup \{\infty\}$. Wir führen nun Maße ein.

1.12 Definition

(a) Sei $\mathcal{R}$ ein Ring in Ω und μ eine Abbildung von $\mathcal{R}$ in $[0, \infty]$ mit $\mu(\emptyset) = 0$. Gilt für jede disjunkte Folge $(A_n)_n$ in $\mathcal{R}$ mit $\bigcup_{n=1}^\infty A_n \in \mathcal{R}$

$$\mu\Big(\bigcup_{n=1}^\infty A_n\Big) = \sum_{n=1}^\infty \mu(A_n), \quad \sigma-\textit{Additivität}, \tag{1.1}$$

so heißt μ ein *Prämaß* (auf $\mathcal{R}$). Gilt (1.1) für jedes endliche System $A_1, \dots, A_n$ von paarweise disjunkten Teilmengen von $\mathcal{R}$, so heißt μ ein *Inhalt*.

(b) Jedes auf einer σ-Algebra $\mathcal{A}$ in Ω definierte Prämaß heißt *Maß* (auf $\mathcal{A}$). $\mu(A)$ für $A \in \mathcal{A}$ wird (μ-)Maß der Menge A genannt. Gilt $\mu(\Omega) < \infty$, heißt das Maß μ *endlich*. $(\Omega, \mathcal{A}, \mu)$ nennt man *Maßraum*. Gilt $\mu(\Omega) = 1$, so heißt μ *Wahrscheinlichkeitsmaß* und $(\Omega, \mathcal{A}, \mu)$ *Wahrscheinlichkeitsraum* (kurz W-Raum). In einem W-Raum heißen die Elemente in $\mathcal{A}$ *Ereignisse*, zu $A \in \mathcal{A}$ heißt $\mu(A)$ *Wahrscheinlichkeit* von A (oder für das Eintreten des Ereignisses A). Elemente ω von Ω mit $\{\omega\} \in \mathcal{A}$ heißen *Elementarereignisse*. Jedes $N \in \mathcal{A}$ mit $\mu(N) = 0$ heißt *μ-Nullmenge*. Die Menge aller μ-Nullmengen bezeichnen wir mit $\mathcal{N}_\mu$.

Wir führen an dieser Stelle bereits die üblichen Sprechweisen und Notationen der Wahrscheinlichkeitstheorie ein. $\emptyset$ bzw. Ω heißen das *unmögliche* bzw. *sichere* Ereignis. Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum (von nun an die übliche Bezeichnung für einen W-Raum). Ereignisse A mit $P(A) = 0$ bzw. $P(A) = 1$ heißen *fast unmöglich* bzw. *fast sicher*.

Falls $A \subset B$, $A, B \in \mathcal{A}$, sagt man, ein Ereignis A *impliziert* B oder *zieht nach sich*. Gilt $A \cap B = \emptyset$, so nennt man A und B *disjunkt*, *fremd* oder *unvereinbar*. Man nennt $A \cup B$ bzw. $A \cap B$ bzw. $A \backslash B$ „mindestens eines der Ereignisse A und B tritt ein“ bzw. „A und B treten ein“ bzw. „es tritt A, nicht aber B ein“. Für eine Folge $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ ist $\bigcup_{n=1}^\infty A_n$ bzw. $\bigcap_{n=1}^\infty A_n$ das Ereignis „A_n tritt für gewisse n ein“ bzw. „A_n tritt ein für alle n“. Schließlich setzen wir

$$\liminf_{n\to\infty} A_n =: \{A_n \text{ für schließlich alle } n\} ,$$
$$\limsup_{n\to\infty} A_n =: \{A_n \text{ für unendlich viele } n\}$$

mit

$$\liminf_{n\to\infty} A_n := \bigcup_{n\geq 1} \bigcap_{m=n}^\infty A_m ,$$
$$\limsup_{n\to\infty} A_n := \bigcap_{n\geq 1} \bigcup_{m=n}^\infty A_m .$$

Man schreibt auch $\{A_n \text{ u.o.}\} := \{A_n \text{ für unendlich viele } n\}$, wobei u.o. „unendlich oft“ bedeutet. Und wir lassen häufig $\{\dots\}$ weg:

$$P(\{A_n \text{ u.o.}\}) = P(\limsup_{n\to\infty} A_n) .$$

1.13 Beispiele

(a) (*Diskrete Maße*, DIRAC-*Maß*) Es sei Ω beliebig und $\tilde\Omega \subset \Omega$ eine höchstens abzählbare Teilmenge. Mit einer Abbildung $p : \tilde\Omega \to (0, \infty]$ definieren wir

$$\mu(A) = \sum_{\omega \in A\cap\tilde\Omega} p(\omega), \quad A \subset \Omega.$$

Dann ist μ ein Maß auf der Potenzmenge von Ω, das auf der diskreten Teilmenge $\tilde{\Omega}$ konzentriert ist (es gilt $\mu(\Omega \setminus \tilde{\Omega}) = 0$). Es gilt dann $\mu(\{\omega\}) = p(\omega) > 0$ für alle $\omega \in \tilde{\Omega}$, und man nennt ω ein *Atom* von μ. Das Maß μ heißt ein *diskretes Maß*. Falls $\sum_{\omega \in \tilde{\Omega}} p(\omega) = 1$, so ist μ ein diskretes Wahrscheinlichkeitsmaß. Falls $\tilde{\Omega}$ einelementig ist, also $\tilde{\Omega} = \{\omega_0\}$ für ein $\omega_0 \in \Omega$, so heißt μ das DIRAC-*Maß* in ω_0, und wir schreiben $\mu = \delta_{\omega_0}$. Es gilt also

$$\delta_{\omega_0}(A) := \begin{cases} 1\,, & \omega_0 \in A \\ 0\,, & \omega_0 \notin A. \end{cases}$$

Ein beliebiges diskretes Maß μ wie oben kann man dann auch als Linearkombination von DIRAC-Maßen auffassen, denn es gilt $\mu = \sum_{\omega \in \tilde{\Omega}} p(\omega)\, \delta_\omega$.

Ist Ω endlich und jedes Ereignis gleichwahrscheinlich, $p(\omega) = \frac{1}{|\Omega|}$, $\omega \in \Omega$, so liegt ein LAPLACE-*Experiment* vor. Hier gilt

$$\mu(A) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \delta_\omega(A) = \frac{|A|}{|\Omega|}.$$

Dies liefert die LAPLACE-*Verteilung* auf Ω. Der n-malige Wurf eines Würfels wird beschrieben durch

$$\Big(\{1, \ldots, 6\}^n\,, \mathcal{P}(\{1, \ldots, 6\}^n)\,, \frac{1}{6^n} \sum_{\omega \in \{1,\ldots,6\}^n} \delta_w\Big)\,.$$

Ein Zufallsexperiment mit nur zwei möglichen Ausgängen heißt BERNOULLI-*Experiment*. Das zugehörige W-Maß ist von der Form

$$\mu = \theta\delta_1 + (1-\theta)\delta_0$$

für ein $\theta \in [0,1]$. μ heißt BERNOULLI-*Verteilung* mit Parameter θ (Münzwurf). Der Fall $\theta = 1/2$ liefert die LAPLACE-Verteilung auf $\{0,1\}$. Jede aus der diskreten Wahrscheinlichkeitstheorie bekannte Verteilung lässt sich auffassen als ein diskretes Maß auf $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$. Wir erinnern an: Die *Binomialverteilung* zu den Parametern n und p,

$$b(n,p) := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k\,; \quad 0 \le p \le 1\,,\ n \in \mathbb{N}\,,$$

ist ein diskretes W-Maß (siehe Abb. 1.1), denn

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1\,.$$

Die POISSON-*Verteilung* zum Parameter $\alpha > 0$,

$$\pi_\alpha := \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} \delta_k\,,$$

ist ein diskretes W-Maß, $\alpha \in \mathbb{R}^+$ (siehe Abb. 1.2), denn $e^\alpha = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}$.

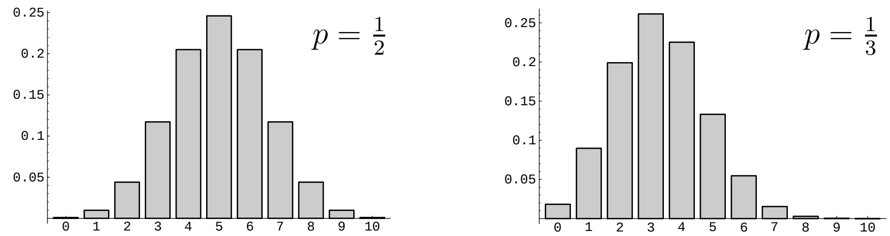


Abb. 1.1: Histogramme der Binomial-Verteilung für $n = 10$.

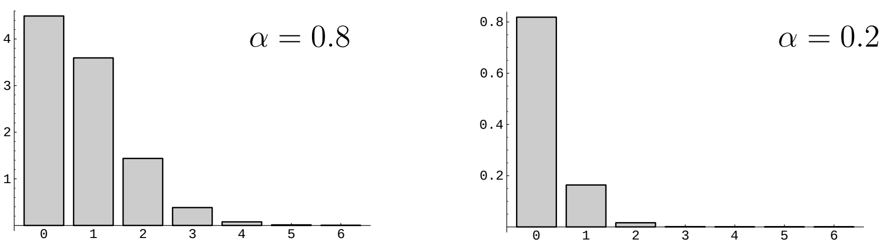


Abb. 1.2: Histogramme der POISSON-Verteilung.

(b) Falls in Beispiel (a) $p(\omega) = 1$ für jedes $\omega \in \tilde{\Omega}$, nennen wir μ das *Zählmaß auf* $\tilde{\Omega}$; dann zählt $\mu(A) = |A \cap \tilde{\Omega}|$ die Anzahl der Punkte aus $\tilde{\Omega}$ in A.

(c) (*Münzwürfe*) Für n Münzwürfe nimmt man

$$\Omega = \big\{(x_1, \ldots, x_n) \,|\, x_i \in \{0,1\}\big\} \, ,$$

für den ∞-fachen

$$\Omega = \{(x_i)_{i \in \mathbb{N}} \,|\, x_i \in \{0,1\}\} = \{0,1\}^{\mathbb{N}} \, .$$

Im einfachen Münzwurf ist $A = \{1\}$ das Ereignis „1 tritt ein“, im n-fachen

$$A = \Big\{(x_1, \ldots, x_n) \,\Big|\, \sum_{i=1}^{n} x_i = k\Big\}$$

„genau k Einsen“ und beim ∞-fachen

$$A = \Big\{(x_i)_{i \in \mathbb{N}} \in \{0,1\}^{\mathbb{N}} \,\Big|\, \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i = p\Big\}$$

„die relative Häufigkeit der 1 ist p“. Setzt man $\mathcal{A} = \sigma(\mathcal{A}_0)$ mit

$$\begin{aligned}\mathcal{A}_0 := \big\{B \subset \{0,1\}^{\mathbb{N}} \,\big|\, &\exists n \in \mathbb{N}\,,\, \exists B_0 \in \mathcal{P}(\{0,1\}^n) \text{ mit} \\ &B = B_0 \times \{0,1\} \times \{0,1\} \times \ldots\big\}\end{aligned}$$

und

$$P\Big(\big\{(x_1, x_2, \dots) \in \{0,1\}^{\mathbb{N}} \,\big|\, x_1 = \bar{x}_1, \dots, x_n = \bar{x}_n\big\}\Big) := 2^{-n}$$

für $\bar{x}_1, \dots, \bar{x}_n \in \{0,1\}$ fest, so gilt:

$$P\left(\left\{(x_i)_{i\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}} \,\middle|\, \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{2}\right\}\right) = 1$$

und obiges P ist fortsetzbar zu einem W-Maß auf $\mathcal{A} = \sigma(\mathcal{A}_0)$. Das beweisen wir etwas später.

(d) Für $A \subset \Omega$ sei $\mu(A) = 0$ für $A = \emptyset$ und $\mu(A) = \infty$ sonst. Dann ist $(\Omega, \mathcal{P}(\Omega), \mu)$ ein Maßraum.

(e) Betrachte $(\mathbb{R}^d, \mathcal{F}^d)$ (siehe Lemma 1.6 und Satz 1.7). Es existiert genau ein Inhalt λ auf $\mathcal{F}^d$, so dass für jedes $I \in \mathcal{R}^d$ der Wert $\lambda(I)$ gleich dem d-dimensionalen Elementarinhalt von I ist. Dabei ist für $I = [a, b)$

$$\lambda(I) := (b_1 - a_1)(b_2 - a_2)\cdots(b_d - a_d)$$

der zugehörige d-dimensionale *Elementarinhalt.*

Beweis: zu (e): Nach 1.6 ist jede Figur $F \in \mathcal{F}^d$ disjunkte endliche Vereinigung von Elementen aus $\mathcal{R}^d : F = I_1 \cup \cdots \cup I_m$. Also folgt für jeden Inhalt λ auf $\mathcal{F}^d$ $\lambda(F) = \lambda(I_1) + \cdots + (I_m)$. Dies bedeutet, dass λ eindeutig durch seine Werte auf $\mathcal{R}^d$ festgelegt ist. Zu zeigen ist somit die *Existenz* von λ. Für $I \in \mathcal{R}^d$ setzen wir $\lambda(I) = d$-dimensionaler Elementarinhalt. λ ist auf $\mathcal{R}^d$ additiv, das heißt für $I_1, \dots, I_m \in \mathcal{R}^d$, paarweise disjunkt, mit $I_1 \cup \cdots \cup I_m \in \mathcal{R}^d$ ist $\lambda(I_1 \cup \cdots \cup I_m) = \lambda(I_1) + \cdots + \lambda(I_m)$ (einfaches Argument). λ wird nun auf $\mathcal{F}^d$ dadurch ausgedehnt, dass jedes $F \in \mathcal{F}^d$ als disjunkte Vereinigung von $I_1, \dots, I_m \in \mathcal{R}^d$ dargestellt werden kann (nicht eindeutig!) und man dann $\lambda(F) := \sum_{j=1}^m \lambda(I_j)$ setzt. Also bleibt zu zeigen: $\lambda(F)$ ist von der Wahl der Darstellung unabhängig. Dazu sei $F = J_1 \cup \cdots \cup J_m = I_1 \cup \cdots \cup I_\ell$ mit $J_1, \dots, J_m, I_1, \dots, I_\ell \in \mathcal{R}^d$. Dann ist

$$J_k = J_k \cap F = \bigcup_{i=1}^{\ell} (J_k \cap I_i),$$

wobei $J_k \cap I_i$ Intervalle sind. Somit ist $\lambda(J_k) = \sum_{i=1}^{\ell} \lambda(J_k \cap I_i)$ und analog $\lambda(I_i) = \sum_{k=1}^{m} \lambda(I_i \cap J_k)$, also

$$\sum_{k=1}^{m} \lambda(J_k) = \sum_{k=1}^{m}\sum_{i=1}^{\ell} \lambda(J_k \cap I_i) = \sum_{i=1}^{\ell}\sum_{k=1}^{m} \lambda(I_i \cap J_k) = \sum_{i=1}^{\ell} \lambda(I_i)\,.$$

Also ist λ auf $\mathcal{F}^d$ endlich additiv und da $\emptyset \in \mathcal{R}^d$ und $\lambda(\emptyset) = 0$, ist λ ein Inhalt. $\square$

Tatsächlich handelt es sich bei λ im Beispiel 1.13 (e) schon um ein Prämaß. Dies ist der Inhalt des folgenden Satzes.

1.14 Satz *λ auf $\mathcal{F}^d$ ist ein Prämaß.*

1.15 Definition Das auf dem Ring $\mathcal{F}^d$ der d-dimensionalen Figuren im $\mathbb{R}^d$ definierte Prämaß λ heißt LEBESGUE*sches Prämaß* im $\mathbb{R}^d$. Es wird mit λ^d bezeichnet. (H. LEBESGUE 1875-1941).

Wir beweisen Satz 1.14 erst, nachdem wir Eigenschaften von Inhalten und Prämaßen kennengelernt haben:

1.16 Satz *Sei μ ein Inhalt auf dem Ring $\mathcal{R}$ und $A, B, (A_n)_n$ seien Mengen aus $\mathcal{R}$. Dann gilt:*

(a) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$

(b) $A \subset B \Rightarrow \mu(A) \leq \mu(B)$

(c) $A \subset B$, $\mu(A) < \infty \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$

(d) $\mu(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mu(A_i)$ *(sub-additiv)*

(e) *Für eine Folge $(A_n)_n$ paarweise fremder Mengen aus $\mathcal{R}$ mit $\bigcup_{n=1}^\infty A_n \in \mathcal{R}$ gilt*

$\sum_{n\geq 1} \mu(A_n) \leq \mu(\bigcup_{n=1}^\infty A_n)$.

Beweis: Es gelten die Identitäten $A\cup B = A\cup(B\setminus A)$ und $B = (A\cap B)\cup(B\setminus A)$, also

$$\mu(A \cup B) = \mu(A) + \mu(B \setminus A)$$

und

$$\mu(B) = \mu(A \cap B) + \mu(B \setminus A).$$

Die Addition dieser beiden Gleichungen liefert (a), wenn $\mu(B \setminus A)$ endlich ist. Ist hingegen $\mu(B \setminus A) = +\infty$, folgt $\mu(A \cup B) = \mu(B) = +\infty$, also auch (a). Ist $A \subset B$, so folgt $\mu(B) = \mu(A) + \mu(B \setminus A)$ womit (b) und (c) gezeigt sind, da $\mu \geq 0$. Sei $B_1 = A_1, B_2 = A_2 \setminus A_1, \ldots, B_n = A_n \setminus (A_1 \cup \cdots \cup A_{n-1})$, so sind $B_1, \ldots, B_n$ disjunkt Mengen aus $\mathcal{R}$, also

$$\mu(\bigcup_{i=1}^n B_i) = \sum_{i=1}^n \mu(B_i).$$

Da $\bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i$ und $B_i \subset A_i$, folgt (d). Ist $A = \bigcup_{n=1}^\infty A_n \in \mathcal{R}$, so gilt $\mu(A_1)+\cdots+\mu(A_m) = \mu(A_1\cup\cdots\cup A_m) \leq \mu(A)$, $m \in \mathbb{N}$ (da disjunkte Mengen), also folgt (e) mit $m \to \infty$. $\square$

1.17 Korollar *Ist μ ein Prämaß auf $\mathcal{R}$, so gilt für $A_0, A_1, \ldots \in \mathcal{R}$: Ist $A_0 \subset \bigcup_{n\geq 1} A_n$ so folgt $\mu(A_0) \leq \sum_{n=1}^{\infty} \mu(A_n)$.*

Beweis: Da $A_0 = \bigcup_{n\geq 1}(A_0 \cap A_n)$ und 1.16 (b) gilt, ist ohne Einschränkung $A_0 = \bigcup_{n\geq 1} A_n$. Setze $B_1 = A_1, \ldots, B_n = A_n \setminus (A_1 \cup \cdots \cup A_{n-1})$ und schließe wie in (d), 1.16. □

Es gilt also $\mu(\bigcup_{n\geq 1} A_n) \leq \sum_{n\geq 1} \mu(A_n)$, falls $\bigcup_{n\geq 1} A_n \in \mathcal{R}$. Dies nennt man die σ-*Subadditivität*.

1.18 Korollar *Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum. Dann ist P σ-subadditiv. Ist I eine endliche Indexmenge, so gilt die Siebformel von* Poincaré-Sylvester*:*

$$\begin{aligned} P\Big(\bigcup_{i\in I} A_i\Big) \quad &= \quad \sum_{\emptyset\neq J\subset I} (-1)^{|J|-1} P\Big(\bigcap_{j\in J} A_j\Big) \\ &\underset{I=\{1,\ldots,n\}}{=} \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\leq i_1<\cdots<i_k\leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}). \end{aligned}$$

Beweis: Die σ-Subadditivität ist klar. Die Siebformel folgt per Induktion aus $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. □

Bezeichnung: Für Mengen $E, E_1, E_2, \ldots$ bezeichne $E_n \uparrow E$ bzw. $E_n \downarrow E$ den Sachverhalt, dass $E_1 \subset E_2 \subset \cdots$, $E = \bigcup_{n\geq 1} E_n$ bzw. $E_1 \supset E_2 \supset \cdots$, $E = \bigcap_{n\geq 1} E_n$ gilt. Man spricht auch von *isotonen* und *antitonen* Folgen.

1.19 Satz *Für einen Inhalt μ auf einem Ring $\mathcal{R}$ betrachte*

(a) *μ ist ein Prämaß.*

(b) *Für jede Folge $(A_n)_n$ von Mengen aus $\mathcal{R}$ mit $A_n \uparrow A$ und $A \in \mathcal{R}$ gilt $\lim_{n\to\infty} \mu(A_n) = \mu(A)$.*

(c) *Für jede Folge $(A_n)_n$ von Mengen aus $\mathcal{R}$ mit $A_n \downarrow A$, $A \in \mathcal{R}$, und $\mu(A_n) < \infty$ für alle $n \in N$ gilt $\lim_{n\to\infty} \mu(A_n) = \mu(A)$.*

(d) *Für jede Folge $(A_n)_n$ von Mengen aus $\mathcal{R}$ mit $A_n \downarrow \emptyset$ und $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$ gilt $\lim_{n\to\infty} \mu(A_n) = 0$.*

Dann gilt:

$$(a) \Leftrightarrow (b) \Rightarrow (c) \Leftrightarrow (d)\,.$$

Ist μ auf $\mathcal{R}$ endlich, gilt also $\mu(A) < +\infty$ für alle $A \in \mathcal{R}$, so sind (a) - (d) äquivalent. (Bezeichnungen: (c) nennt man Stetigkeit von oben, (b) entsprechend Stetigkeit von unten).

Beweis: $(a) \Rightarrow (b)$: Es sei $A_0 := \emptyset$ und $B_n := A_n \setminus A_{n-1}$ für $n \in \mathbb{N}$. Dann ist $(B_n)_n$ eine Folge paarweise disjunkter Mengen in $\mathcal{R}$ und $A = \bigcup_{n\geq 1} B_n$ und $A_n = B_1 \cup \cdots \cup B_n$. Damit ist $\mu(A) = \sum_{n\geq 1} \mu(B_n) = \lim_{n\to\infty} \sum_{i=1}^n \mu(B_i) = \lim_{n\to\infty} \mu(A_n)$, da μ σ-additiv ist.

$(b) \Rightarrow (a)$: Sei $(A_n)_n$ eine Folge disjunkter Mengen aus $\mathcal{R}$ mit $A := \bigcup_{n\geq 1} A_n \in \mathcal{R}$. Mit $B_n := A_1 \cup \cdots \cup A_n$ ist $B_n \uparrow A$, also $\mu(A) = \lim_{n\to\infty} \mu(B_n)$ und $\mu(B_n) = \sum_{i=1}^n \mu(A_i)$, also $\mu(A) = \sum_{n\geq 1} \mu(A_n)$, also ist μ ein Prämaß.

$(b) \Rightarrow (c)$: Aus $A_n \downarrow A$ folgt $A_1 \setminus A_n \uparrow A_1 \setminus A$ (alle Mengen liegen in $\mathcal{R}$). Also ist $\mu(A_1 \setminus A) = \lim_{n\to\infty} \mu(A_1 \setminus A_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n)$ nach 1.16 (c): da $A \subset A_n$, folgt $\mu(A) < \infty$ und daher $\mu(A_1 \setminus A) = \mu(A_1) - \mu(A)$. Also folgt (c).

$(c) \Rightarrow (d)$ ist klar

$(d) \Rightarrow (c)$: Aus $A_n \downarrow A$ folgt $A_n \setminus A \downarrow \emptyset$. Da $A_n \setminus A \subset A_n$, ist $\mu(A_n \setminus A)$ endlich und somit $\lim_{n\to\infty} \mu(A_n \setminus A) = 0$. Da $\mu(A) \leq \mu(A_n) < +\infty$, folgt mit $\mu(A_n \setminus A) = \mu(A_n) - \mu(A)$ die Behauptung.

Sei nun μ endlich. In diesem Fall zeigen wir $(d) \Rightarrow (b)$: Es sei $(A_n)_n$ eine Folge in $\mathcal{R}$ mit $A_n \uparrow A \in \mathcal{R}$, dann ist $A \setminus A_n \downarrow \emptyset$, also $0 = \lim_{n\to\infty} \mu(A \setminus A_n) = \lim_{n\to\infty} \big(\mu(A) - \mu(A_n)\big)$, da μ endlich. $\square$

1.20 Bemerkung Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum. Für P sind alle Aussagen in Satz 1.19 äquivalent. Mittels Satz 1.19 kann man auf eine andere Art und Weise die σ-Subadditivität von P sehen:

$$P\Big(\bigcup_{i=1}^\infty A_i\Big) = \lim_{n\to\infty} P\Big(\bigcup_{i=1}^n A_i\Big) \leq \lim_{n\to\infty} \sum_{i=1}^n P(A_i) = \sum_{i=1}^\infty P(A_i),$$

wobei die erste Gleichheit aus 1.19 folgt und die Ungleichung aus $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$.

Ein einfaches, aber für die W-Theorie sehr wichtiges Resultat können wir jetzt bereits zeigen:

1.21 Lemma (von Borel-Cantelli) *Es seien $A_i \in \mathcal{A}$, $i \in \mathbb{N}$. Dann gilt:*

$$\sum_{i=1}^\infty P(A_i) < \infty \ \Rightarrow\ P(\limsup_{n\to\infty} A_n) = P\Big(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m\Big) = 0\,.$$

Beweis: Da $\bigcup_{m\geq n} A_m \downarrow \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m$, folgt

$$P(\limsup_{n\to\infty} A_n) \overset{1.19}{=} \lim_{n\to\infty} P\Big(\bigcup_{m\geq n} A_m\Big) \leq \lim_{n\to\infty} \sum_{m=n}^\infty P(A_m) = 0$$

nach Vorraussetzung, wobei die letzte Ungleichung die σ-Subadditivität verwendet. □

Beweis von Satz 1.14: Da λ auf $\mathcal{F}^d$ endlich ist, zeigen wir (d) in Satz 1.19. Es sei $(F_n)_n$ eine Folge in $\mathcal{F}^d$ mit $F_n \downarrow \emptyset$. Dann ist zu zeigen: $\lim_{n\to\infty} \lambda(F_n) = 0$. Angenommen $\lim_{n\to\infty} \lambda(F_n) = \inf_{n\in\mathbb{N}} \lambda(F_n) =: \delta > 0$. Jedes F_n ist endliche Vereinigung disjunkter Intervalle. Dann existiert ein $G_n \in \mathcal{F}^d$ mit $\bar{G}_n \subset F_n$ und

$$\lambda(F_n) - \lambda(G_n) \leq \frac{\delta}{2^n}\,.$$

Es sei $H_n := G_1 \cap \cdots \cap G_n$. Dann ist $(H_n)_n$ eine Folge aus $\mathcal{F}^d$ mit $H_n \supset H_{n+1}$, $n \in \mathbb{N}$, und $\bar{H}_n \subset \bar{G}_n \subset F_n$. Da F_n beschränkt ist, ist $\bar{H}_n$ kompakt (Satz aus der Analysis) und $F_n \supset \bar{H}_n \supset \bar{H}_{n+1}$, also $\bigcap_{n=1}^\infty \bar{H}_n \neq \emptyset$, falls $H_n \neq \emptyset$ für jedes $n \in \mathbb{N}$ (dies ist eine Aussage über kompakte Mengen, die wir als Übung lassen). Dann wäre aber $\bigcap_{n=1}^\infty F_n \neq \emptyset$, im Widerspruch zur Annahme. Es bleibt also noch zu zeigen: für $n \in \mathbb{N}$ ist $H_n \neq \emptyset$. Dazu zeigen wir

$$\lambda(H_n) \geq \lambda(F_n) - \delta(1 - 2^{-n}) \qquad (*).$$

Dies genügt, denn mit $(*)$ ist $\lambda(H_n) \geq \delta - \delta(1 - 2^{-n}) = \delta/2^n > 0$, $n \in \mathbb{N}$. Wir zeigen $(*)$ via Induktion: Der Fall $n = 1$ folgt so: $H_1 = G_1$ und $\lambda(F_1) - \lambda(G_1) \leq 2^{-1}\delta$ nach Konstruktion. Nun folgt der Induktionsschritt $n \to n+1$: $H_{n+1} = G_{n+1} \cap H_n$, also $\lambda(H_{n+1}) = \lambda(G_{n+1}) + \lambda(H_n) - \lambda(G_{n+1} \cup H_n)$ (wir verwenden 1.16 (a)). Nach der Induktionsannahme für $\lambda(H_n)$ und der Wahl von G_{n+1} folgt $\lambda(H_{n+1}) \geq \lambda(F_{n+1}) - 2^{-n-1}\,\delta + \lambda(F_n) - \delta(1-2^{-n}) - \lambda(F_n) = \lambda(F_{n+1}) - \delta(1-2^{-n-1})$, da $G_{n+1} \cup H_n \subset F_{n+1} \cup F_n = F_n$. Somit ist alles gezeigt. □

KAPITEL 2

Fortsetzung von Prämaßen zu Maßen

In diesem Kapitel wollen wir das LEBESGUEsche Prämaß λ^d fortsetzen zu einem Maß, welches auf der σ-Algebra $\sigma(\mathcal{F}^d)$ definiert ist. Wir konstruieren dazu zuerst sogenannte äußere Maße, die auf allen Teilmengen einer gegebenen Menge definiert sind, und einige, aber nicht alle Eigenschaften eines Maßes besitzen. Anschließend schränken wir das äußere Maß auf geeignete Teilmengen der Potenzmenge ein. Eine auf CARATHÉODORY zurückgehende geschickte Auswahl dieser Teilmengen liefert dann einen geeigneten Maßraum. Genauer werden wir ein Prämaß μ auf einem Ring $\mathcal{R}$ betrachten und es zu einem äußeren Maß μ^* ausdehnen. Dann liefert μ^* eingeschränkt auf die von $\mathcal{R}$ erzeugte σ-Algebra ein Maß, welches auf $\mathcal{R}$ mit dem Prämaß μ übereinstimmt. Wir klären weiter die Frage nach der Eindeutigkeit einer solchen Fortsetzung und betrachten die σ-Algebra der BORELschen Mengen $\sigma(\mathcal{F}^d)$.

Ω sei stets eine nichtleere Menge.

2.1 Definition Eine Abbildung $\mu^* : \mathcal{P}(\Omega) \to [0,\infty]$ mit $\mu^*(\emptyset) = 0$ heißt *äußeres Maß* auf Ω, wenn sie wachsend und σ-subadditiv ist, also mit $A_1 \subset A_2$ stets $\mu^*(A_1) \leq \mu^*(A_2)$ folgt und

$$\mu^*\Big(\bigcup_{n\geq 1} A_n\Big) \leq \sum_{n\geq 1} \mu^*(A_n)$$

für $(A_n)_n$ in $\mathcal{P}(\Omega)$ gilt.

2.2 Bemerkungen

(a) Ein äußeres Maß ist stets auf ganz $\mathcal{P}(\Omega)$ definiert.

(b) Jedes Maß auf $\mathcal{P}(\Omega)$ ist ein äußeres Maß (siehe Korollar 1.17).

(c) Es sei $A \subset \Omega$ und

$$\mu^*(A) := \begin{cases} 0\,, & A = \emptyset \\ 1\,, & A \neq \emptyset \end{cases},$$

dann ist μ^* ein äußeres Maß auf Ω (und es ist genau dann ein Maß, wenn Ω einpunktig ist).

Einem Prämaß μ auf einem Ring $\mathcal{R}$ können wir durch die folgende Konstruktion ein äußeres Maß zuordnen:

2.3 Satz *Sei μ ein Prämaß auf einem Ring $\mathcal{R}$ in Ω. Sei $\mu^* : \mathcal{P}(\Omega) \to [0,\infty]$ definiert durch*

$$\mu^*(Q) := \inf\Big\{\sum_{n\geq 1}\mu(A_n), (A_n)_n \in \mathcal{R}^{\mathbb{N}}, Q \subset \bigcup_{n\geq 1} A_n\Big\},$$

mit $Q \in \mathcal{P}(\Omega)$ (setze $\inf\emptyset := \infty$). Dann ist μ^* ein äußeres Maß auf Ω (das von $(\mathcal{R},\mu)$ induzierte äußere Maß) mit $\mu^*(A) = \mu(A)$ für alle $A \in \mathcal{R}$.

2.4 Beispiel $\Omega = \mathbb{R}^d$, $\mathcal{R} = \mathcal{F}^d$, $\mu = \lambda^d$. Dann heißt $(\lambda^d)^*$ das d-dimensionale Lebesgue*sche äußere Maß*.

Beweis: Die Eigenschaften $\mu^*(\emptyset) = 0$ und μ^* ist wachsend sind nach Definition klar. Wir zeigen die σ-Subadditivität: ohne Einschränkung sei $\mu^*(Q_n) < \infty$, $n \in \mathbb{N}$, $Q_n \in \mathcal{P}(\Omega)$. Zu $\epsilon > 0$ und $n \in \mathbb{N}$ existiert eine Folge $(A_{nm})_m \in \mathcal{R}^{\mathbb{N}}$ mit $Q_n \subset \bigcup_{m\geq 1} A_{nm}$ und

$$\sum_{m\geq 1}\mu(A_{nm}) \leq \mu^*(Q_n) + \frac{\epsilon}{2^n}\ .$$

Dann ist $\bigcup_{n\geq 1} Q_n \subset \bigcup_{n,m\geq 1} A_{nm}$, $A_{nm} \in \mathcal{R}$, $n,m \in \mathbb{N}$, also

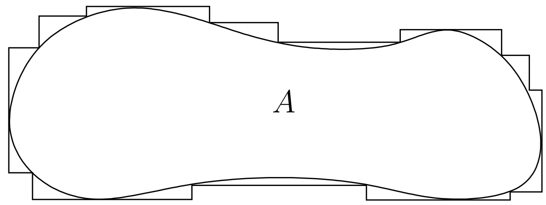
$$\begin{aligned}\mu^*(\bigcup_{n\geq 1} Q_n) &\leq \sum_{n,m\geq 1}\mu(A_{nm}) \leq \sum_{n\geq 1}\big(\mu^*(Q_n) + \frac{\epsilon}{2^n}\big)\\ &= \sum_{n\geq 1}\mu^*(Q_n) + \epsilon\ .\end{aligned}$$

Weiter ist $\mu^* \geq 0$ (klar).

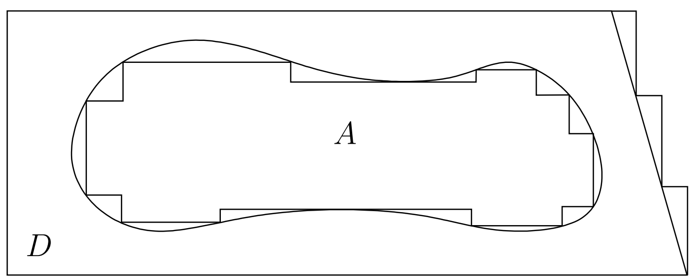
Aus Korollar 1.17 folgt $\mu(A) \leq \mu^*(A)$, $A \in \mathcal{R}$. Mit $A_1 := A$, $A_n := \emptyset$, $n \geq 2$, ist $A \subset \bigcup_{n\geq 1} A_n$ und somit $\mu^*(A) \leq \sum_{n\geq 1}\mu(A_n) = \mu(A)$. Also $\mu(A) = \mu^*(A)$ für $A \in \mathcal{R}$. □

2.5 Bemerkung Ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$ enthalte die leere Menge und eine Folge $(A_n)_n$ mit $\Omega = \bigcup_{n\geq 1} A_n$. Eine Abbildung $\mu : \mathcal{A} \to [0,\infty]$ erfülle $\mu(\emptyset) = 0$. Dann ist μ^*, definiert wie in Satz 2.3 für jedes $Q \subset \Omega$ ein äußeres Maß auf Ω (das von $(\mathcal{A},\mu)$ induzierte). Hierbei muß man nicht $\inf\emptyset := 0$ setzen. Der Beweis geht völlig analog. Diese systematische Konstruktion dient zur Herleitung Lebesgue-Stieltjesscher äußerer Maße und Hausdorffscher äußerer Maße. Wir kommen auf die erst genannten zurück.

Für eine Teilmenge $A \subset \mathbb{R}^d$ ist das äußere Maß $(\lambda^d)^*(A)$ eine Approximation des Inhalts von A *von außen* durch Figuren. A wird von außen durch Figuren approximiert.



Anstelle von A betrachten wir nun $D \setminus A$, wobei D eine beschränkte Obermenge D von A in $\mathbb{R}^d$ ist. $D \setminus A$ werde nun analog von außen durch $(\lambda^d)^*(D \setminus A)$ approximiert.



Dann bedeutet dies eine Approximation von A *von innen* durch Figuren. Man könnte daher das *innere Maß* von A relativ zu D durch

$$(\lambda^d)^*(D) - (\lambda^d)^*(D \setminus A)$$

definieren. Es ist dann zu erwarten, dass diejenigen Teilmengen A von $\mathbb{R}^d$ eine ausgezeichnete Rolle spielen, deren äußeres Maß mit dem inneren Maß bezüglich jeder beschränkten Obermenge übereinstimmt:

$$(\lambda^d)^*(D) = (\lambda^d)^*(A) + (\lambda^d)^*(D \setminus A), \quad D \subset \mathbb{R}^d, D \supset A.$$

Dies führt für ein beliebiges äußeres Maß zu der folgenden Definition.

2.6 Definition Es sei μ^* ein äußeres Maß auf Ω. $A \subset \Omega$ heißt μ^*-*messbar*, wenn für jedes $D \subset \Omega$

$$\mu^*(D) \geq \mu^*(A \cap D) + \mu^*(A^c \cap D)$$

gilt ($\leq$ gilt sowieso). Die Menge aller μ^*-messbaren Teilmengen von Ω bezeichnen wir mit $\mathcal{A}(\mu^*)$.

2.7 Satz (von Carathéodory) *Es sei μ^* ein äußeres Maß auf Ω. Dann ist $\mathcal{A}(\mu^*)$ eine σ-Algebra auf Ω und $\mu := \mu^*|\mathcal{A}(\mu^*)$ ein Maß auf $\mathcal{A}(\mu^*)$. Hierbei bezeichnet $\mu^*|\mathcal{A}(\mu^*)$ die Restriktion von μ^* auf $\mathcal{A}(\mu^*)$.*

Beweis: Es gilt $\emptyset \in \mathcal{A}(\mu^*)$ und aus $A \in \mathcal{A}(\mu^*)$ folgt $A^c \in \mathcal{A}(\mu^*)$ nach Definition der μ^*-messbaren Mengen. Seien $A, B \in \mathcal{A}(\mu^*)$ und $D \subset \Omega$. Dann ist $\mu^*(D) \geq \mu^*(A \cap D) + \mu^*(A^c \cap D)$ und $\mu^*(A^c \cap D) \geq \mu^*(B \cap A^c \cap D) + \mu^*(B^c \cap A^c \cap D)$. Da μ^* subadditiv ist, folgt daraus

$$\mu^*(D) \geq \mu^*\big((A \cap D) \cup (B \cap A^c \cap D)\big) + \mu^*(B^c \cap A^c \cap D).$$

Nun gilt $\big(A \cup (B \cap A^c)\big) \cap D = (A \cup B) \cap D$, also folgt

$$\mu^*(D) \geq \mu^*\big((A \cup B) \cap D\big) + \mu^*\big((A \cup B)^c \cap D\big),$$

also ist $A \cup B$ μ^*-messbar. Wir haben also bereits gezeigt, dass $\mathcal{A}(\mu^*)$ eine Algebra ist. Sei nun $(A_j)_j$ eine disjunkte Folge in $\mathcal{A}(\mu^*)$. Somit ist wegen der μ^*-Messbarkeit von A_1

$$\mu^*((A_1 \cup A_2) \cap D) = \mu^*\big(((A_1 \cup A_2) \cap D) \cap A_1\big) + \mu^*\big(((A_1 \cup A_2) \cap D) \cap A_1^c\big),$$

also

$$\mu^*((A_1 \cup A_2) \cap D) = \mu^*(A_1 \cap D) + \mu^*(A_2 \cap D) \ ,$$

da die Mengen A_1 und A_2 disjunkt sind. Mittels vollständiger Induktion folgt somit

$$\mu^*\left(\left(\bigcup_{j=1}^{m} A_j\right) \cap D\right) = \sum_{j=1}^{m} \mu^*(A_j \cap D), \ m \in \mathbb{N} \ . \tag{2.1}$$

Mit $A := \bigcup_{j \geq 1} A_j$ folgt mit der Monotonie von μ^*

$$\mu^*(A \cap D) \geq \sum_{j=1}^{m} \mu^*(A_j \cap D), \ m \in \mathbb{N} \ .$$

Für $m \to \infty$ folgt also

$$\mu^*(A \cap D) \geq \sum_{j \geq 1} \mu^*(A_j \cap D) \ .$$

Die σ-Subadditivität liefert also insgesamt

$$\mu^*(A \cap D) = \sum_{j=1}^{\infty} \mu^*(A_j \cap D) \ . \tag{2.2}$$

Da wir schon gesehen hatten, dass $\mathcal{A}(\mu^*)$ eine Algebra ist, gilt

$$\mu^*(D) = \mu^*\left(\left(\bigcup_{j=1}^{m} A_j\right)^c \cap D\right) + \mu^*\left(\left(\bigcup_{j=1}^{m} A_j\right) \cap D\right).$$

Mittels der Monotonie und (2.1) folgt

$$\mu^*(D) \geq \mu^*(A^c \cap D) + \sum_{j=1}^{m} \mu^*(A_j \cap D).$$

Jetzt folgt mit $m \to \infty$ und (2.2) die Ungleichung $\mu^*(D) \geq \mu^*(A^c \cap D) + \mu^*(A \cap D)$, also ist A μ^*-meßbar. Somit ist nach 1.4(c) $\mathcal{A}(\mu^*)$ eine σ-Algebra. Setze abschließend in (2.2) $D = \Omega$, so folgt, dass $\mu^*|\mathcal{A}(\mu^*)$ ein Maß auf $\mathcal{A}(\mu^*)$ ist. □

2.8 Satz (Fortsetzungssatz) *Es sei μ ein Prämaß auf einem Ring $\mathcal{R}$ in Ω und μ^* das zugehörige äußere Maß auf Ω (siehe Satz 2.3). Dann gilt: $\sigma(\mathcal{R}) \subset \mathcal{A}(\mu^*)$ und $\bar{\mu} := \mu^*|\sigma(\mathcal{R})$ ist ein Maß auf $\sigma(\mathcal{R})$, welches μ erweitert.*

Beweis: Für $A \in \mathcal{R}$ wollen wir zeigen, dass $\mu^*(D) \geq \mu^*(D \cap A) + \mu^*(D \cap A^c)$ für jedes $D \subset \Omega$. Es sei $(A_n)_n \in \mathcal{R}^{\mathbb{N}}$ mit $D \subset \bigcup_{n\geq 1} A_n$, so ist

$$\begin{aligned}
\sum_{n\geq 1} \mu(A_n) \quad &= \quad \sum_{n\geq 1} \mu(A_n \cap A) + \sum_{n\geq 1} \mu(A_n \cap A^c) \\
&\overset{(2.3)}{=} \quad \sum_{n\geq 1} \mu^*(A_n \cap A) + \sum_{n\geq 1} \mu^*(A_n \cap A^c) \\
&\overset{\sigma-subadd.}{\geq} \quad \mu^*\left(\bigcup_{n\geq 1} A_n \cap A\right) + \mu^*\left(\bigcup_{n\geq 1} A_n \cap A^c\right) \\
&\overset{monoton}{\geq} \quad \mu^*(D \cap A) + \mu^*(D \cap A^c)\ .
\end{aligned}$$

Man betrachte nun das Infimum über alle solche Überdeckungen von D. Dann folgt $\mathcal{R} \subset \mathcal{A}(\mu^*)$. $\mathcal{A}(\mu^*)$ ist nach Satz 2.7 eine σ-Algebra, also folgt $\sigma(\mathcal{R}) \subset \mathcal{A}(\mu^*)$. Da $\mu^*|\mathcal{A}(\mu^*)$ nach Satz 2.7 ein Maß ist, ist auch $\mu^*|\sigma(\mathcal{R})$ ein Maß und nach Satz 2.3 gilt $\mu^*|\mathcal{R} = \mu$. Dies war zu zeigen. □

Es ergibt sich zusammenfassend das folgende Schema:

$$\begin{array}{ccccccc}
\mathcal{R} & \subset & \sigma(\mathcal{R}) & \subset & \mathcal{A}(\mu^*) & \subset & \mathcal{P}(\Omega) \\
 & & & & & & \\
\underbrace{\mu}_{\text{input}} & \leftarrow & \underbrace{\mu^*|\sigma(\mathcal{R})}_{\text{output}} & \leftarrow & \mu^*|\mathcal{A}(\mu^*) & \leftarrow & \mu^*
\end{array}$$

2.9 Satz (Eindeutigkeitssatz) *Stimmen zwei Maße μ und ν, die auf einer σ-Algebra $\mathcal{A}$ definiert sind, auf einem durchschnittsstabilen Erzeuger $\mathcal{C}$ von $\mathcal{A}$ überein, und existiert eine Folge $\Omega_n \in \mathcal{C}$, $n \in \mathbb{N}$, mit $\Omega_n \nearrow \Omega$ und $\mu(\Omega_n) = \nu(\Omega_n) < \infty$, so gilt $\mu = \nu$ auf $\mathcal{A}$.*

Beweis: Dies ist ein erster Beweis, in dem das Dynkin-System-Argument verwendet wird. Mit der Stetigkeit der Maße μ und ν wissen wir

$$\mu(\Omega) = \lim_{n\to\infty} \mu(\Omega_n) = \lim_{n\to\infty} \nu(\Omega_n) = \nu(\Omega).$$

Sei zunächst $\mu(\Omega) = \nu(\Omega) < \infty$. Es gilt ja $\mu(\Omega) = \lim_{n\to\infty} \mu(\Omega_n) = \lim_{n\to\infty} \nu(\Omega_n) = \nu(\Omega)$. Wir zeigen:

$$\mathcal{D} := \{A \in \mathcal{A} : \mu(A) = \nu(A)\} \text{ ist ein Dynkin-System,}$$

denn dann folgt $\mathcal{A} = \mathcal{D}$. Es gilt $\Omega \in \mathcal{D}$. Ist $D \in \mathcal{D}$, so ist $\mu(D^c) = \mu(\Omega) - \mu(D) = \nu(\Omega) - \nu(D) = \nu(D^c)$, also $D^c \in \mathcal{D}$. Für jede Folge $(D_n)_n$ von paarweise disjunkten Mengen aus $\mathcal{D}$ gilt

$$\mu\Big(\bigcup_{n=1}^{\infty} D_n\Big) = \sum_{n=1}^{\infty} \mu(D_n) = \sum_{n=1}^{\infty} \nu(D_n) = \nu\Big(\bigcup_{n=1}^{\infty} D_n\Big)\ ,$$

also $\bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$.

Für den allgemeinen Fall sei μ_n, ν_n definiert durch

$$\mu_n(A) := \mu(A \cap \Omega_n),\ \nu_n(A) := \nu(A \cap \Omega_n),\ A \in \mathcal{A}\ .$$

Es gilt $\mu_n = \nu_n$ für alle $n \in \mathbb{N}$. Es folgt für alle $A \in \mathcal{A}$

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n) = \lim_{n\to\infty} \nu(A \cap \Omega_n) = \nu(A)\ ,$$

also $\mu = \nu$. □

Um eine Bedingung für die Eindeutigkeit des Maßes $\bar{\mu}$ in Satz 2.8 angeben zu können, definieren wir:

2.10 Definition Ein Prämaß μ auf einem Ring $\mathcal{R}$ auf Ω heißt σ-*endlich*, falls es $(A_n)_n \in \mathcal{R}^{\mathbb{N}}$ mit $\Omega = \bigcup_{n\geq 1} A_n$ und $\mu(A_n) < \infty,\ n \in \mathbb{N}$, gibt.

2.11 Beispiel Das Lebesguesche Prämaß λ^d (Definition 1.15) ist σ-endlich, denn $[-n, n) \subset \mathbb{R}^d$ hat die Eigenschaft $[-n, n) \uparrow \mathbb{R}^d$ und $\lambda^d([-n, n)) < \infty$.

Zusammenfassend erhalten wir

2.12 Satz *Es sei μ ein Prämaß auf einem Ring $\mathcal{R}$ in Ω und μ^* das zugehörige äußere Maß auf Ω. Ist μ σ-endlich, so ist $\mu^*|\sigma(\mathcal{R})$ das einzige Maß, welches μ erweitert.*

Beweis: Nur die Eindeutigkeit von $\mu^*|\sigma(\mathcal{R})$ ist zu zeigen. Diese folgt aus Satz 2.9, denn mit der σ-Endlichkeit von μ besitzt der Ring $\mathcal{R}$ alle Eigenschaften des in Satz 2.9 auftretenden Erzeugers $\mathcal{C}$. □

2.13 Definition Zu $(\mathbb{R}^d, \mathcal{F}^d)$ heißt die σ-Algebra $\sigma(\mathcal{F}^d) = \sigma(\mathcal{R}^d)$ die σ-Algebra der BOREL*schen Mengen* von $\mathbb{R}^d$, in Zeichen $\mathcal{B}^d := \sigma(\mathcal{F}^d)$. Das von $(\lambda^d)^*$ auf $\mathbb{R}^d$ induzierte Maß $(\lambda^d)^*|\mathcal{B}^d$ heißt LEBESGUE-BOREL*sches Maß* oder auch d-dimensionales LEBESGUEmaß auf $\mathbb{R}^d$ und wird mit λ^d bezeichnet (vgl. Definition 1.15). Nach Satz 2.12 ist es das einzige Maß auf $\mathcal{B}^d$, welches jedem nach rechts halboffenen Intervall in $\mathbb{R}^d$ seinen d-dimensionalen Elementarinhalt zuordnet. Im Fall $d = 1$ schreiben wir $\mathbb{R}$ bzw. $\mathcal{B}$.

Naheliegende Fragen zu $\mathcal{B}^d$ sind: Welche Mengen sind in $\mathcal{B}^d$ enthalten? Gibt es in $\mathcal{P}(\mathbb{R}^d)$ überhaupt nicht-BORELsche Mengen?

Wir betrachten zunächst:

2.14 Lemma *Jede offene Teilmenge O in $\mathbb{R}^d$ kann als Vereinigung einer Folge $(I_n)_n$ von Intervallen der Form $[a, b)$, $a, b \in \mathbb{Q}^d$, dargestellt werden ($(I_n)_n \in (\mathcal{R}^d)^{\mathbb{N}}$).*

Beweis: O kann als Vereinigung abzählbar vieler offener, beschränkter Intervalle dargestellt werden (Intervalle mit lauter rationalen Koordinaten der Eckpunkte). Jedes beschränkte Intervall der Form $(a, b) := \{x \in \mathbb{R}^d : a < x < b\}$ ist Vereinigung einer Folge von Intervallen aus $\mathcal{R}^d$: $[\bar{a}_n, b) \uparrow (a, b)$ mit $\bar{a}_n := (\min(a_1 + \frac{1}{n}, b_1), \ldots, \min(a_d + \frac{1}{n}, b_d))$. Die Vereinigung einer Folge abzählbarer Mengen ist wieder abzählbar. Also ist die Behauptung bewiesen. □

Man kann sogar zeigen, dass in Lemma 2.14 eine disjunkte Folge $(I_n)_n$ gefunden werden kann.

2.15 Satz *Es bezeichne $\mathcal{O}^d$ bzw. $\mathcal{C}^d$ bzw. $\mathcal{K}^d$ das System aller offenen bzw. abgeschlossenen bzw. kompakten Teilmengen von $\mathbb{R}^d$. Dann ist*

$$\mathcal{B}^d = \sigma(\mathcal{O}^d) = \sigma(\mathcal{C}^d) = \sigma(\mathcal{K}^d).$$

Beweis: Da jede kompakte Menge abgeschlossen ist, folgt $\mathcal{K}^d \subset \mathcal{C}^d \subset \sigma(\mathcal{C}^d)$, also folgt $\sigma(\mathcal{K}^d) \subset \sigma(\mathcal{C}^d)$. Sei $K_n := \bar{B}(0, n)$ die kompakte Vollkugel vom

Radius n. Dann ist $C \in \mathcal{C}^d$ darstellbar als $C = \bigcup_{n\geq 1}(C \cap K_n)$ und $C \cap K_n$ ist kompakt, also $\mathcal{C}^d \subset \sigma(\mathcal{K}^d)$, also $\sigma(\mathcal{C}^d) = \sigma(K^d)$. Offene Mengen sind Komplemente der abgeschlossenen Mengen, also folgt $\sigma(\mathcal{C}^d) = \sigma(\mathcal{O}^d)$. Damit bleibt zu zeigen: $\sigma(\mathcal{O}^d) = \mathcal{B}^d$: jedes $[a, b[\in \mathcal{R}^d$ ist Durchschnitt einer Folge offener und beschränkter Intervalle: $(\bar{a}_n, b) \downarrow [a, b)$ mit $\bar{a}_n := (a_1 - \frac{1}{n}, \dots, a_d - \frac{1}{n})$, also gilt $\mathcal{R}^d \subset \sigma(\mathcal{O}^d)$ und somit $\mathcal{B}^d = \sigma(\mathcal{R}^d) \subset \sigma(\mathcal{O}^d)$. Mit Hilfe von Lemma 2.14 folgt $\mathcal{O}^d \subset \sigma(\mathcal{R}^d) = \mathcal{B}^d$, also $\sigma(\mathcal{O}^d) \subset \mathcal{B}^d$ und somit $\sigma(\mathcal{O}^d) = \mathcal{B}^d$.
□

2.16 Beispiele

(a) Jede BORELsche Menge in $\mathbb{R}^d$, die Teilmenge einer *Koordinatenhyperebene* ist, ist eine LEBESGUE-BORELsche Nullmenge. Ohne Einschränkung seien $H := \mathbb{R}^{d-1} \times \{0\}$, $\epsilon > 0$ und $k \in \mathbb{N}$. Setze $\epsilon_k := \epsilon \cdot (2k)^{-d+1} \cdot 2^{-(k+2)}$ und $J_k(\epsilon) := [-k, k)^{d-1} \times [-\epsilon_k, \epsilon_k)$, dann ist $H \subset \bigcup_{k\geq 1} J_k(\epsilon)$ und $\lambda^d(J_k(\epsilon)) = \epsilon \cdot 2^{-(k+1)}$ und somit $\sum_{k\geq 1} \lambda^d(J_k(\epsilon)) = \frac{\epsilon}{2} < \epsilon$.

Wir werden später sehen: Jede BORELsche Teilmenge von $\mathbb{R}^d$, die in einem affinen echten Unterraum enthalten ist, ist eine LEBESGUE-BORELsche Nullmenge.

(b) Jede abzählbare BORELsche Teilmenge des $\mathbb{R}^d$ ist eine LEBESGUE-BORELsche Nullmenge. Wir wissen, dass $\{x\} \subset \mathbb{R}^d$ abgeschlossen ist und daher BORELsch. Weiter findet man ein H wie in (a) mit $\{x\} \subset H$. Dann liefert die σ-Additivität die Behauptung.

(c) $F \subset \mathbb{R}$ sei CANTORs Diskontinuum (siehe Analysis). Es ist $\lambda^1(F) = 0$.

(d) Für $a, b \in \mathbb{R}^d$ mit $a \leq b$ gilt $\lambda^d([a, b)) = \lambda^d((a, b)) = \lambda^d([a, b]) = \lambda^d((a, b])$. Dies folgt einfach mittels der aufsteigenden bzw. absteigenden Intervall-Folgen, wie sie im Beweis von Lemma 2.14 und Satz 2.15 vorkommen.

Ausgangspunkt der Konstruktion des LEBESGUE-Maßes λ^1 auf $(\mathbb{R}, \mathcal{B})$ bildete die Setzung $\lambda([a, b)) := b - a$ für nach rechts halboffene Intervalle. Wir wollen nun $b - a$ durch $F(b) - F(a)$ für ein monotones $F : \mathbb{R} \to \mathbb{R}$ ersetzen. Unter welchen Zusatzeigenschaften an F liefert dies ein Maß auf $(\mathbb{R}, \mathcal{B})$?

2.17 Definition Es sei $F : \mathbb{R} \to \mathbb{R}$ monoton wachsend und linksseitig stetig. Dann heißt F *maßerzeugende Funktion*. Gelten außerdem $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to\infty} F(x) = 1$, so nennt man F *Verteilungsfunktion*.

2.18 Satz *Eine Funktion $F : \mathbb{R} \to \mathbb{R}$ definiert genau dann durch $\mu_F([a, b)) := F(b) - F(a)$ für alle $a, b \in \mathbb{R}$, $a \leq b$, und eindeutige Fortsetzung ein σ-endliches Maß μ_F auf $(\mathbb{R}, \mathcal{B})$, wenn sie maßerzeugend ist.*

Beweis: μ_F sei ein Maß auf $(\mathbb{R}, \mathcal{B})$, welches $\mu_F([a,b)) = F(b) - F(a)$ erfüllt. Da $\mu_F([a,b)) \geq 0$ für alle $a \leq b$, folgt die Monotonie von F. Weiter gilt für alle $a \in \mathbb{R}$ und jede Folge $(a_n)_n$ in $\mathbb{R}$ mit $a_n \uparrow a$, dass $[a_1, a_n) \uparrow [a_1, a)$. Dann folgt mit Satz 1.19

$$\begin{aligned}\lim_{n\to\infty} F(a_n) - F(a_1) &= \lim_{n\to\infty} \mu_F([a_1, a_n)) \\ &= \mu_F([a_1, a)) = F(a) - F(a_1)\ ,\end{aligned}$$

also die linksseitige Stetigkeit in a.

Die Rückrichtung verläuft analog zu der Konstuktion des eindimensionalen LEBESGUE-Maßes λ: Zu F existiert genau ein Inhalt μ auf dem Ring $\mathcal{F}$ der 1-dimensionalen Figuren mit der Eigenschaft $\mu([a,b)) = F(b) - F(a)$ (analog zu Beispiele 1.13(e)). Man benötigt nur die Monotonie von F. Da F linksseitig stetig ist, gibt es zu jedem $[a,b) \in \mathcal{R}$ und zu jedem $\varepsilon > 0$ ein $[a,c) \in \mathcal{R}$ mit $\overline{[a,c)} = [a,c] \subset [a,b]$ und

$$\mu([a,b)) - \mu([a,c)) = \mu([c,b)) = F(b) - F(c) \leq \varepsilon\ .$$

Dann aber folgt wie in Satz 1.14, dass μ ein σ-endliches Prämaß auf $\mathcal{F}$ ist. Dies kann nach dem Satz von CARATHÉODORY zu einem Maß $\tilde{\mu}$ auf $\mathcal{B}$ fortgesetzt werden. Dieses Maß leistet das Gewünschte. □

2.19 Definition Ist μ ein W-Maß auf $(\mathbb{R}, \mathcal{B})$, so heißt $F_\mu(x) := \mu\big([-\infty, x)\big)$ *Verteilungsfunktion von* μ. Das äußere Maß μ_F^* nennt man das von F erzeugte LEBESGUE-STIELTJES*sche äußere Maß*. Ist $F : \mathbb{R} \to \mathbb{R}$ eine maßerzeugende Funktion, so wird das von μ_F^* auf $\mathbb{R}$ erzeugte Maß als das von F induzierte LEBESGUE-STIELTJES*sche Maß* auf $\mathbb{R}$ bezeichnet. Wir schreiben dafür μ_F.

2.20 Bemerkung Ersetzt man linksseitig stetig in der Definition einer maßerzeugenden Funktion durch rechtsseitig stetig, so bleibt der Satz 2.18 richtig, wenn man alle linksseitig abgeschlossenen Intervalle durch rechtsseitig abgeschlossene substituiert.

2.21 Beispiele

(a) Die Verteilungsfunktion zur POISSON-Verteilung ist

$$F_{\pi_\alpha}(x) = \begin{cases} e^{-\alpha} \sum_{k=0}^{n} \frac{\alpha^k}{k!} & \text{, falls } n < x \leq n+1 \\ 0 & \text{, falls } x < 0\ . \end{cases}$$

(b) Sei sei $f : \mathbb{R} \to \mathbb{R}^+$ eine stetige nichtnegative Funktion, so dass das RIEMANN-Integral $\int_a^b f(x)\, dx$ endlich ist für alle $a, b \in \mathbb{R}$ mit $a < b$. Dann

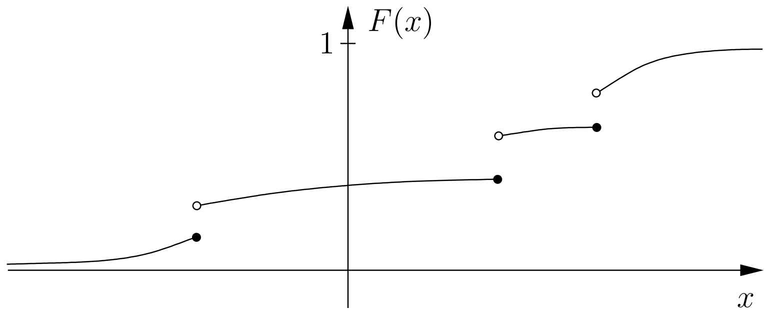


Abb. 2.1: Eine Verteilungsfunktion.

existiert genau ein Maß μ_f auf $(\mathbb{R}, \mathcal{B})$ mit

$$\mu_f([a,b)) = \int_a^b f(x)\,dx$$

für alle $a < b$. Man sagt, μ_f habe eine *Dichte*. Wir wenden also Satz 2.18 auf $F(t) = \int_0^t f(x)\,dx$ an. Dieses Beispiel kann man stark verallgemeinern, siehe ein späteres Kapitel.

KAPITEL 3

Messbare Abbildungen und das Lebesgue-Borel-Maß

In Analogie zu stetigen Abbildungen zwischen metrischen Räumen, bei denen die Urbilder der offenen Mengen offen sind, betrachten wir in diesem Kapitel Abbildungen, die die Urbilder von Elementen einer σ-Algebra in eine σ-Algebra führen. Diese Abbildungen heißen messbare Abbildungen. Mit Hilfe messbarer Abbildungen können auch Maße abgebildet werden. Weiter betrachten wir Abbildungseigenschaften des LEBESGUE-BORELschen Maßes λ^d. Der Flächeninhalt einer messbaren ebenen Punktmenge (also einer Menge in $\mathcal{B}^2$) ändert sich nicht, wenn man die Menge einer beliebigen Drehung oder Verschiebung unterwirft. Dies nennt man die Bewegungsinvarianz des LEBESGUE-BORELschen Maßes. Wir untersuchen hier sogar das Verhalten von λ^d bei beliebigen invertierbaren affinen Abbildungen. Weiter zeigen wir die Existenz nicht BORELscher Mengen in $\mathbb{R}^d$.

3.1 Definition Es seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ Messräume (siehe Definition 1.1) und $T\colon \Omega \to \Omega'$ eine Abbildung. T heißt $\mathcal{A}/\mathcal{A}'$-*messbar*, wenn gilt: $T^{-1}(A') \in \mathcal{A}$ für alle $A' \in \mathcal{A}'$ ($T^{-1}(\mathcal{A}') \subset \mathcal{A}$). Die $\mathcal{A}/\mathcal{A}'$-Messbarkeit drücken wir symbolisch auch durch

$$T : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$$

aus. Ist $\Omega' = \mathbb{R}^d$, $\mathcal{A}' = \mathcal{B}^d$, so heißt eine $\mathcal{A}/\mathcal{B}^d$-messbare Abbildung BOREL-*messbar*. Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $(\Omega', \mathcal{A}')$ ein Messraum. Dann heißt eine messbare Abbildung $X : (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$ *Zufallsvariable*, im Fall $(\Omega', \mathcal{A}') \subset (\mathbb{R}, \mathcal{B})$ *Zufallsgröße* und im Fall $(\Omega', \mathcal{A}') \subset (\mathbb{R}^d, \mathcal{B}^d)$ für $d \geq 2$ *Zufallsvektor* ($X = (X_1, \ldots, X_d)$). Wenn es im Zusammenhang klar ist, welche σ-Algebren gegeben sind und keine Verwechslung möglich ist, sagen wir fortan auch kurz *messbar*.

3.2 Beispiele

(a) Jede konstante Abbildung $T\colon \Omega \to \Omega'$ ist $\mathcal{A}/\mathcal{A}'$-messbar.

(b) Es sei $(\Omega, \mathcal{A})$ ein Messraum und zu $A \subset \Omega$ sei

$$1_A(\omega) := \begin{cases} 1\,, & \omega \in A \\ 0\,, & \omega \in A^c \end{cases} .$$

Diese Abbildung heißt *Indikatorfunktion* oder *charakteristische Funktion* von A. 1_A ist genau dann $\mathcal{A}/\mathcal{B}$-messbar, wenn $A \in \mathcal{A}$ gilt (denn $1_A^{-1}(B)$, $B \subset \mathbb{R}$, liegt in $\{\Omega, A, A^c, \emptyset\}$).

3.3 Satz *Es seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ Messräume und $\mathcal{E}'$ sei ein Erzeuger von $\mathcal{A}'$. Dann ist $T\colon \Omega \to \Omega'$ genau dann $\mathcal{A}/\mathcal{A}'$-messbar, wenn $T^{-1}(E') \in \mathcal{A}$ für alle $E' \in \mathcal{E}'$ gilt.*

Beweis: Die Menge aller $\mathcal{Q}' := \{Q' \subset \Omega' \colon T^{-1}(Q') \in \mathcal{A}\}$ ist eine σ-Algebra in Ω' (einfache Übung). $\mathcal{A}' \subset \mathcal{Q}'$ ist gleichbedeutend mit der Messbarkeit von T. $\mathcal{A}' \subset \mathcal{Q}'$ ist gleichbedeutend mit $\mathcal{E}' \subset \mathcal{Q}'$ und $\mathcal{E}' \subset \mathcal{Q}'$ ist gleichbedeutend mit $T^{-1}(E') \in \mathcal{A}$ für alle $E' \in \mathcal{E}'$. □

3.4 Beispiel Jede stetige Abbildung $T\colon \mathbb{R}^d \to \mathbb{R}^\ell$ ist $\mathcal{B}^d/\mathcal{B}^\ell$-messbar (Borel-messbar). Dies folgt, da $\mathcal{O}^\ell$ ein Erzeuger von $\mathcal{B}^\ell$ ist (siehe Satz 2.15) und $T^{-1}(O) \in \mathcal{O}^d \subset \mathcal{B}^d$ für alle $O \in \mathcal{O}^\ell$ wegen der Stetigkeit. Also folgt die Behauptung aus Satz 3.3.

3.5 Satz *Gegeben seien drei Messräume $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$ und $(\Omega_3, \mathcal{A}_3)$. Sind $T_1\colon (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ und $T_2\colon (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$ messbare Abbildungen, so ist $T_2 \circ T_1$ $\mathcal{A}_1/\mathcal{A}_3$-messbar.*

Beweis: $(T_2 \circ T_1)^{-1}(A) = T_1^{-1}(T_2^{-1}(A))$. □

3.6 Definition Es sei eine Familie von Messräumen $((\Omega_i, \mathcal{A}_i))_{i \in I}$ und eine Familie $(T_i)_{i \in I}$ von Abbildungen $T_i : \Omega \to \Omega_i$ einer Menge Ω in die einzelnen Mengen Ω_i gegeben. Dann ist die von $\cup_{i \in I} T_i^{-1}(\mathcal{A}_i)$ in Ω erzeugte σ-Algebra die kleinste σ-Algebra $\mathcal{A}$, bezüglich welcher jedes T_i $\mathcal{A}/\mathcal{A}_i$-messbar ist. Wir bezeichnen diese σ-Algebra mit $\sigma(T_i; i \in I)$, also

$$\sigma(T_i; i \in I) := \sigma\Big(\bigcup_{i \in I} T_i^{-1}(\mathcal{A}_i)\Big).$$

Sie heißt die *von den Abbildungen T_i erzeugte σ-Algebra*. Im Fall einer endlichen Menge $I = \{1, \dots, n\}$ schreiben wir $\sigma(T_1, \dots, T_n)$. Offenbar gilt $\sigma(T_1) = T_1^{-1}(\mathcal{A}_1)$.

Mit Hilfe messbarer Abbildungen können Maße abgebildet werden:

3.7 Satz *Es sei $T\colon (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ eine messbare Abbildung. Dann wird für jedes Maß μ auf $(\Omega, \mathcal{A})$ durch $A' \mapsto \mu(T^{-1}(A'))$ ein Maß μ' auf $\mathcal{A}'$ definiert.*

Beweis: Mit $(A'_n)_n$, disjunkte Folge in $\mathcal{A}'$, ist $(T^{-1}(A'_n))_n$ disjunkte Folge in $\mathcal{A}$ und es gilt

$$T^{-1}\big(\bigcup_{n \geq 1} A'_n\big) = \bigcup_{n \geq 1} T^{-1}(A'_n).$$

Daraus folgt schon die Behauptung. □

3.8 Definition Das Maß μ' in Satz 3.7 heißt *Bildmaß von μ unter T* und wird mit $T(\mu)$ bezeichnet:

$$T(\mu)(A') := \mu(T^{-1}(A')), \quad A' \in \mathcal{A}'.$$

Wir schreiben auch μT^{-1} oder μ^T. Ist $X : (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$ eine Zufallsvariable, so heißt das Bildmaß von P unter X *Verteilung von X* (*unter P*) und wird mit $P^X = X(P) = PX^{-1}$ bezeichnet.

3.9 Bemerkung Es sei die Situation von Satz 3.5 gegeben. Weiter sei μ ein Maß auf $(\Omega_1, \mathcal{A}_1)$. Dann ist

$$(T_2 \circ T_1)(\mu) = T_2(T_1(\mu)) \qquad \textit{(Transitivität)}$$

Denn für $A \in \mathcal{A}_3$ ist

$$\begin{aligned} \mu\big((T_2 \circ T_1)^{-1}(A)\big) &= \mu\big(T_1^{-1}(T_2^{-1}(A))\big) \\ &= T_1(\mu)\left(T_2^{-1}(A)\right) = T_2(T_1(\mu))(A). \end{aligned}$$

3.10 Beispiele Es seien $(\Omega, \mathcal{A}) = (\Omega', \mathcal{A}') = (\mathbb{R}^d, \mathcal{B}^d)$ und $\mu = \lambda^d$.

(a) Zu $a \in \mathbb{R}^d$ sei $T_a\colon \mathbb{R}^d \to \mathbb{R}^d$ definiert durch $T_a(x) := x + a$. T_a ist stetig, also messbar nach Beispiel 3.4. Wie sieht $T_a(\lambda^d)$ aus? Es ist $T_a^{-1} = T_{-a}$. Für $[b, c) \in \mathcal{R}^d$ ist

$$T_a(\lambda^d)([b,c)) = \lambda^d(T_a^{-1}([b,c))) = \lambda^d(T_{-a}([b,c))) = \lambda^d([b-a, c-a)) = \lambda^d([b,c)),$$

also stimmen λ^d und $T_a(\lambda^d)$ auf $\mathcal{R}^d$ überein. Somit folgt mit Satz 2.9 (siehe auch Definition 2.13) $\lambda^d = T_a(\lambda^d)$. Man spricht von der *Translationsinvarianz* von λ^d.

(b) Zu $r \in \mathbb{R}\backslash\{0\}$ und $p \in \{1, \dots, d\}$ sei $M_{p,r}\colon \mathbb{R}^d \to \mathbb{R}^d$ die lineare Abbildung

$$M_{p,r}(x_1, \dots, x_d) = \begin{array}{c} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & 0 & \\ & & 1 & & & & \\ & & & r & & & \\ & & & & 1 & & \\ & 0 & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \\ p\text{-te Stelle} \end{array} \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$$

(p-te Koordinate wird durch rx_p ersetzt). Jede lineare Abbildung ist stetig und somit BOREL-messbar. Für $[a,b) \in \mathcal{R}^d$ ist das Urbild unter $M_{p,r}$ gleich $[a',b')$ für $r > 0$ bzw. $(b',a']$ für $r < 0$, wobei $a' = (a_1,\ldots,a_{p-1},\frac{a_p}{r},a_{p+1},\ldots,a_d)$ und b' analog. Also ist

$$\lambda^d\big(M_{p,r}^{-1}([a,b))\big) = \frac{1}{|r|}\lambda^d([a,b)) \ .$$

Somit stimmen $M_{p,r}(\lambda^d)$ und $\frac{1}{|r|}\lambda^d$ auf $\mathcal{R}^d$ überein und mit Satz 2.9 folgt

$$M_{p,r}(\lambda^d) = \frac{1}{|r|}\lambda^d = \frac{1}{|\det M_{p,r}|}\lambda^d.$$

(c) Betrachte zu $a \in \mathbb{R}^d$, $a_i \neq 0$, $i = 1,\ldots,d$, die Abbildung $M_a := M_{1,a_1} \circ M_{2,a_2} \circ \cdots \circ M_{d,a_d}$, also $M_a(x_1,\ldots,x_d) = (a_1x_1,\ldots,a_dx_d)$. Dann ist

$$M_a(\lambda^d) = \frac{1}{|a_1 \ldots a_d|}\lambda^d = \frac{1}{|\det M_a|}\lambda^d$$

nach Bemerkung 3.9 und Beispiel (b).

Im Spezialfall $a = (r,\ldots,r) \in \mathbb{R}^d$ nennt man die Abbildung M_a eine *Homothetie*. Der Spezialfall einer Homothetie mit $r = -1$ ist die Spiegelung am Nullpunkt.

Das LEBESGUE-BORELsche-Maß λ^d auf $\mathcal{B}^d$ ist durch die Translationsinvarianz und eine Normierung bereits eindeutig bestimmt:

3.11 Satz *Ist μ ein translationsinvariantes Maß auf $\mathcal{B}^d$ mit $\mu([0,1)) = 1$, so ist $\mu = \lambda^d$.*

Beweis: Für $n_1,\ldots,n_d \in \mathbb{N}$ betrachte man das Gitter der Punkte $(\frac{k_1}{n_1},\ldots,\frac{k_d}{n_d})$, $0 < k_j \leq n_j$, $k_j \in \mathbb{N}$, $j = 1,\ldots,d$. Verschiebt man das Intervall $[0,\frac{1}{n_1}) \times \cdots \times [0,\frac{1}{n_d})$ um jeden dieser Gitterpunkte, so folgt die disjunkte Zerlegung

$$[0,1) = \bigcup_{\substack{0<k_j\leq n_j\\ j=1,\ldots,d}} \left(\prod_{i=1}^{d} [0,\frac{1}{n_i}) + \left(\frac{k_1}{n_1},\ldots,\frac{k_d}{n_d}\right)\right).$$

Alle $n_1 \cdots n_d$ Mengen rechts haben gleiches Maß μ (da μ translationsinvariant ist) und mit $\mu([0,1)) = 1$ folgt $\mu\left(\prod_{j=1}^d [0,\frac{1}{n_j})\right) = \frac{1}{n_1\cdots n_d}$. Wendet man nochmals die Translationsinvarianz von μ an, so folgt

$$\mu([a,b)) = \lambda^d([a,b)), \quad a,b \in \mathbb{Q}^d.$$

Aber Intervalle $[a,b)$ mit $a,b \in \mathbb{Q}^d$ bilden einen Ring $\mathcal{R}_\mathbb{Q}^d$ (klar nach Satz 1.7). Mit Satz 2.9 ist $\mu = \lambda^d$ auf $\sigma(\mathcal{R}_\mathbb{Q}^d)$. Nun ist $\sigma(\mathcal{R}_\mathbb{Q}^d) \subset \mathcal{B}^d$ und nach Lemma 2.14 ist $\mathcal{O}^d \subset \sigma(\mathcal{R}_\mathbb{Q}^d)$, also $\mathcal{B}^d = \sigma(\mathcal{O}^d) \subset \sigma(\mathcal{R}_\mathbb{Q}^d)$ und somit $\mathcal{B}^d = \sigma(\mathcal{R}_\mathbb{Q}^d)$. Damit ist der Satz bewiesen. □

3.12 Korollar

(a) *Ist μ ein translationsinvariantes Maß auf $\mathcal{B}^d$ mit $\mu([0,1]) = 1$, so ist $\mu = \lambda^d$.*

(b) *Ist μ ein translationsinvariantes Maß auf $\mathcal{B}^d$ mit $\alpha = \mu([0,1])$, so ist $\mu = \alpha\lambda^d$.*

Beweis: zu (a): Es gilt $1 = \mu([0,1]) \le \mu([0,2)) = 2^d\alpha$ mit $\alpha := \mu([0,1))$, also ist $\alpha > 0$. Nun erfüllt $\nu := \frac{1}{\alpha}\mu$ die Voraussetzungen von 3.11, also $\nu = \lambda^d$. Damit ist

$$\frac{1}{\alpha} = \nu([0,1]) = \lambda^d([0,1]) = 1,$$

also folgt $\mu = \lambda^d$.

zu (b): Für $\alpha > 0$ erfüllt $\alpha^{-1}\mu$ die Voraussetzungen von (a). Für $\alpha = 0$ ist $\mu([0,1]) = 0$ und

$$\mu(\mathbb{R}^d) = \mu\Big(\bigcup_{g\in\mathbb{Z}^d} ([0,1) + g)\Big) = \sum_{g\in\mathbb{Z}^d} \mu([0,1)) = 0,$$

und somit ist $\mu = 0$. □

Die Beispiele 3.10 zeigen das Verhalten von λ^d unter speziellen linearen Abbildungen. Dies kann verallgemeinert werden:

3.13 Satz *Es sei $f\colon \mathbb{R}^d \to \mathbb{R}^d$ eine bijektive affine Abbildung. Dann ist f* Borel*-messbar und*

$$f(\lambda^d) = |\det f|^{-1}\lambda^d.$$

Dabei bedeutet $\det f$ *die Determinante der f darstellenden Matrix.*

3.14 Korollar *Es sei $f\colon \mathbb{R}^d \to \mathbb{R}^d$ eine bijektive affine Abbildung. Dann ist für jedes $A \in \mathcal{B}^d$ auch $f(A) \in \mathcal{B}^d$ und*

$$\lambda^d(f(A)) = |\det f|\,\lambda^d(A).$$

Beweis: Mit f ist f^{-1} bijektiv und affin. Man wende 3.13 auf f^{-1} statt f an. □

Es sei eine affine Abbildung $f\colon \mathbb{R}^d \to \mathbb{R}^d$ der Form $f(x) = g(x) + a$ gegeben, wobei $a \in \mathbb{R}^d$ und $g\colon \mathbb{R}^d \to \mathbb{R}^d$ eine orthogonale Abbildung ist. Ein solches f heißt *Bewegung*. f ist genau dann eine Bewegung, wenn für alle $x, y \in \mathbb{R}^d$ $\|f(x) - f(y)\| = \|x - y\|$ gilt (siehe Lineare Algebra). Für jede Bewegung f ist $|\det f| = 1$. Somit folgt das

3.15 Korollar (Bewegungsinvarianz des Lebesgue-Borel-Maßes)
Jede Bewegung ist Borel*-messbar und es gilt* $f(\lambda^d) = \lambda^d$, $\lambda^d(f(A)) = \lambda^d(A)$, $A \in \mathcal{B}^d$.

Satz 3.13 gilt sogar für C^1-Diffeomorphismen. Dies ist die Transformationsformel aus der Analysis III. Wir verweisen daher auf dieses Resultat. Doch wir sind mit Hilfe eines Arguments aus der Linearen Algebra schon so dicht am direkten Beweis, das wir ihn hier ausführen.

Beweis von Satz 3.13: Da f stetig ist, ist f auch Borel-messbar nach Beispiel 3.4. Da λ^d translationsinvariant ist, genügt es mit Bemerkung 3.9, den Satz für Abbildungen $g\colon \mathbb{R}^d \to \mathbb{R}^d$ zu betrachten, die linear und bijektiv sind, für die also $\det g \neq 0$ gilt. Es sei also $g \in \mathrm{GL}(d, \mathbb{R})$, wobei $\mathrm{GL}(d, \mathbb{R})$ die allgemeine lineare Gruppe bezeichnet. Es sei weiter $A \in \mathcal{B}^d$ und $a \in \mathbb{R}^d$, so gilt

$$T_a\big(g(\lambda^d)\big)(A) = g(\lambda^d)(A - a) = \lambda^d\big(g^{-1}(A) - g^{-1}(a)\big) = \lambda^d\big(g^{-1}(A)\big) = g(\lambda^d)(A).$$

Also ist $g(\lambda^d)$ translationsinvariant. Die Menge $g^{-1}\big([0,1]\big)$ ist kompakt, es folgt also $g(\lambda^d)([0,1]) < \infty$. Mit Korollar 3.12(b) und $c(g) := g(\lambda^d)([0,1])$ schließen wir

$$g(\lambda^d) = c(g)\,\lambda^d.$$

Damit bleibt zu zeigen:

$$c(g) = |\det g|^{-1}. \tag{3.1}$$

Dies behandeln wir für unterschiedliche g.

(a) Ist g orthogonal, so ist

$$c(g)\,\lambda^d\big(B(0,1)\big) = g(\lambda^d)\big(B(0,1)\big) = \lambda^d\big(g^{-1}(B(0,1)\big) = \lambda^d\big(B(0,1)\big).$$

Also ist $c(g) = 1 = |\det g|^{-1}$, und somit ist (3.1) für orthogonale g bewiesen.

(b) $g \in \mathrm{GL}(d, \mathbb{R})$ sei bezüglich der kanonischen Basis durch eine Diagonalmatrix beschrieben. Dann liefert Beispiel 3.10 (c) $c(g) = |\det g|^{-1}$.

(c) Jetzt kommt ein Fakt aus der Linearen Algebra: $g \in \mathrm{GL}(d, \mathbb{R})$ hat eine Darstellung $g = v\,d\,w$ (im Sinne der Komposition von Abbildungen) mit orthogonalen Abbildungen v und w und d ist eine durch eine Diagonalmatrix beschriebene Abbildung wie unter (b). Dies sieht man mit

Hilfe des Äquivalenz-Satzes für positiv definite Matrizen (Koecher, Lineare Algebra und analytische Geometrie, S.195) wie folgt: bezeichnet g^* den adjungierten Endomorphismus von g, so ist $g\,g^*$ positiv definit. Zu dieser Abbildung gibt es nach dem Äquivalenz-Satz eine orthogonale Abbildung v und eine diagonale Abbildung d, so dass $g\,g^* = v\,d^2\,v^*$ gilt. Die Abbildung $w := d^{-1}\,v^*\,g$ ist orthogonal und es gilt $g = v\,d\,w$. Dann gilt $|\det g| = |\det d|$, und Bemerkung 3.9 und die Fälle (a) und (b) liefern die Behauptung. □

Abschließend betrachten wir die Existenz nicht-Borelscher Mengen.

3.16 Satz *Es gilt $\mathcal{B}^d \neq \mathcal{P}(\mathbb{R}^d)$ für jedes $d \in \mathbb{N}$.*

Beweis: Wir erklären auf $\mathbb{R}^d$ eine Äquivalenzrelation: zu $x, y \in \mathbb{R}^d$ sei $x \sim y \Leftrightarrow x - y \in \mathbb{Q}^d$. Die Äquivalenzklassen haben dann die Form $x + \mathbb{Q}^d, x \in \mathbb{R}^d$. Zu jeder reellen Zahl η existiert ein $n \in \mathbb{Z}$ mit $n \leq \eta < n + 1$, also $\eta - n \in [0, 1)$. Somit gibt es in jeder Äquivalenzklasse ein $x \in [0, 1)$. Es gibt nach dem Auswahlaxiom eine Menge $K \subset [0, 1)$, so dass K mit jeder Äquivalenzklasse genau ein Element gemeinsam hat. Also ist

$$\mathbb{R}^d = \bigcup_{k \in K} (k + \mathbb{Q}^d) = \bigcup_{y \in \mathbb{Q}^d} (y + K).$$

Weiter gilt für $y_1 \neq y_2 \Rightarrow (y_1 + K) \cap (y_2 + K) = \emptyset$, denn ist der Durchschnitt nicht-leer, so existieren $k_1, k_2 \in K$ mit $y_1 + k_1 = y_2 + k_2$, also $k_1 \sim k_2$. Dann ist $k_1 = k_2$, also $y_1 = y_2$ im Widerspruch zu $y_1 \neq y_2$. Falls nun $K \in \mathcal{B}^d$, so folgt aus der σ-Additivität

$$\sum_{y \in \mathbb{Q}^d} \lambda^d(y + K) = \lambda^d(\mathbb{R}^d) = +\infty.$$

Nach 3.10 (a) (Translationsinvarianz) ist $\lambda^d(y + K) = \lambda^d(K)$ für $y \in \mathbb{Q}^d$. Also muß $\lambda^d(K) > 0$ sein. Mit $K \subset [0, 1)$ ist

$$\bigcup_{y \in [0,1) \cap \mathbb{Q}^d} (y + K) \subset [0, 2)$$

und daher

$$\sum_{y \in [0,1) \cap \mathbb{Q}^d} \lambda^d(y + K) \leq \lambda^d([0, 2)) = 2^d < \infty.$$

Das geht aber nur, wenn $\lambda^d(y + K) = \lambda^d(K) = 0$, im Widerspruch zu $\lambda^d(K) > 0$. Also kann K nicht in $\mathcal{B}^d$ liegen. □

KAPITEL 4

Messbare numerische Funktionen

Der Begriff *messbare* Abbildungen erschien im Hinblick auf die Definition von *Bildmaßen* als genau der richtige Begriff. Wir werden nun sehen, dass die Konstruktion des Integrals den Begriff der Messbarkeit in ein neues Licht bringt. Es seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $A \in \mathcal{A}$. Das Integral von 1_A über Ω bezüglich des Maßes μ sollte durch $\mu(A)$ festgelegt sein. Also muß A zu $\mathcal{A}$ gehören. Diese Verträglichkeit von 1_A mit der σ-Algebra $\mathcal{A}$ wird für kompliziertere Funktionen $f : \Omega \to \mathbb{R}$ durch ein geeignetes Approximationsargument verallgemeinert und führt zurück zum Begriff der Messbarkeit, gegeben in der Definition 3.1.

4.1 Definition Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$. Eine Funktion $f : \Omega \to \overline{\mathbb{R}}$ heißt *numerisch*. f heißt $\mathcal{A}$-messbare numerische Funktion, falls $f^{-1}(-\infty)$, $f^{-1}(\infty)$ und $f^{-1}(O)$ für jede offene Teilmenge O in $\mathbb{R}$ zu $\mathcal{A}$ gehören. Die Menge aller $\mathcal{A}$-messbaren numerischen Funktionen auf Ω bezeichnen wir mit

$$\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}) =: \mathcal{L}_0.$$

4.2 Bemerkung Es bezeichne $\overline{\mathcal{B}}$ das System aller Untermengen B von $\overline{\mathbb{R}}$ mit $B \cap \mathbb{R} \in \mathcal{B}$. Man nennt $\overline{\mathcal{B}}$ die BORELsche σ-Algebra auf $\overline{\mathbb{R}}$. Es ist f eine $\mathcal{A}$-messbare numerische Funktion genau dann, wenn f $\mathcal{A}/\overline{\mathcal{B}}$-messbar ist (siehe Definition 3.1 und Satz 3.3).

Von nun an verwenden wir die folgenden Bezeichnungen: Sind $f, g : \Omega \to \overline{\mathbb{R}}$ und $\alpha, \beta \in \overline{\mathbb{R}}$, so setze

$$[f > \alpha] := [\alpha < f] := \{x \in \Omega : f(x) > \alpha\} = f^{-1}\big((\alpha, \infty]\big).$$

Analog $[f < \alpha]$, $[f \leq \alpha]$, $[f \geq \alpha]$, $[f = \alpha]$, $[f \neq \alpha]$, $[\alpha < f \leq \beta]$, $[f < g]$, $[f \leq g]$, $[f \neq g]$, $[f = g]$, $[\alpha < f, g > \beta]$ usw.

4.3 Satz *Für eine numerische Funktion $f : \Omega \to \overline{\mathbb{R}}$ sind äquivalent.*

(a) $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$

(b) $[f < \alpha] \in \mathcal{A}$ *für jedes* $\alpha \in \mathbb{Q}$ *[bzw.* $\mathbb{R}$*]*

(c) $[f \leq \alpha] \in \mathcal{A}$ *für jedes* $\alpha \in \mathbb{Q}$ *[bzw.* $\mathbb{R}$*]*

(d) $[f > \alpha] \in \mathcal{A}$ *für jedes* $\alpha \in \mathbb{Q}$ *[bzw.* $\mathbb{R}$*]*

(e) $[f \geq \alpha] \in \mathcal{A}$ *für jedes* $\alpha \in \mathbb{Q}$ *[bzw.* $\mathbb{R}$*]*.

Beweis: (a) $\Rightarrow$ (b): Nach Voraussetzung liegen $f^{-1}(-\infty)$, $f^{-1}\big((-\infty,\alpha)\big)$, $\alpha \in \mathbb{Q}$ (bzw. $\mathbb{R}$) in $\mathcal{A}$. Es gilt

$$[f<\alpha] = f^{-1}\big([-\infty,\alpha)\big) = f^{-1}(-\infty) \cup f^{-1}\big((-\infty,\alpha)\big),$$

also folgt $[f<\alpha]\in\mathcal{A}$.
Weiter gelten die folgenden Identitäten:

$$[f\le\alpha] = \bigcap_{j=1}^{\infty}[f<\alpha+(1/j)],\ [f>\alpha]=[f\le\alpha]^c,\ [f\ge\alpha]=\bigcap_{j=1}^{\infty}[f>\alpha-(1/j)].$$

Damit ist bereits (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) bewiesen.

(e) $\Rightarrow$ (a): Es sei O offen in $\mathbb{R}$. Nach Lemma 2.14 existieren $(\alpha_j)_j$ und $(\beta_j)_j$ in $\mathbb{Q}^{\mathbb{N}}$ mit $O=\bigcup_{j\ge1}[\alpha_j,\beta_j)$, also

$$f^{-1}(O) = \bigcup_{j\ge1} f^{-1}\big([\alpha_j,\beta_j)\big) = \bigcup_{j\ge1}\big([f\ge\alpha_j]\cap[f<\beta_j]\big).$$

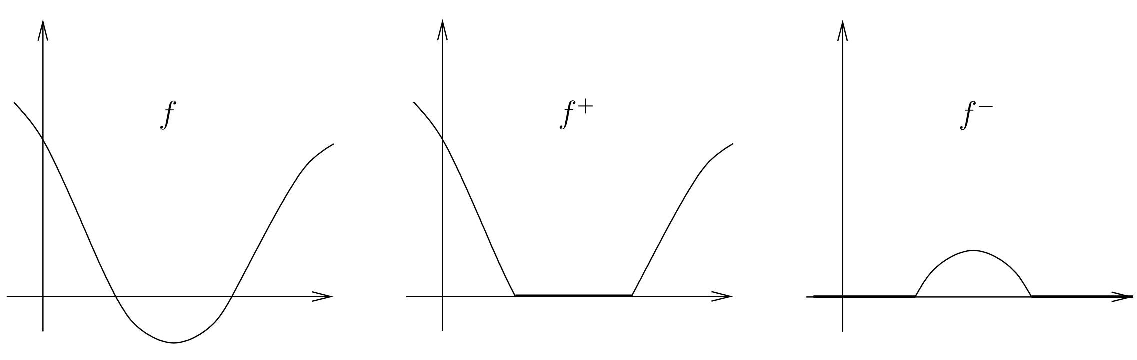
Mit $[f<\beta_j]=[f\ge\beta_j]^c$ folgt somit aus (e) $f^{-1}(O)\in\mathcal{A}$. Weiter gilt

$$f^{-1}(-\infty)=\bigcap_{j\ge1}[f<-j],\quad f^{-1}(\infty)=\bigcap_{j\ge1}[f\ge j],$$

und somit liegt auch $f^{-1}(\pm\infty)$ in $\mathcal{A}$. $\square$

4.4 Definition Es sei $f:\Omega\to\overline{\mathbb{R}}$. Dann heißen $f^+ := f\vee 0$ bzw. $f^- := 0\vee(-f)$ *Positiv* - bzw. *Negativteil* von f. (Erinnerung: $\vee$ steht für das Supremum).

In der folgenden Abbildung sind der Positiv- und der Negativteil eines $f\in\mathbb{R}^\Omega$ dargestellt.



Es gelten

$$f^+\ge0,\ f^-\ge0,\ f=f^+-f^-,\ |f|=f^++f^-.$$

Diese Identitäten sind anschaulich klar. Für ihren formalen Beweis (Übung) verwendet man die Rechenregel

$$(f \vee g) + h = (f + h) \vee (g + h), \quad (-f) \vee (-g) = -(f \wedge g)$$

für $f, g, h \in \overline{\mathbb{R}}^{\Omega}$, wobei $\wedge$ für das Infimum steht.

4.5 Satz *Es seien $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ und $(f_j)_j$ eine Folge in $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ und $k \in \mathbb{N}$. Dann gehören die numerischen Funktionen f^+, f^-, $|f|$, $\max_{1 \le j \le k} f_j$, $\min_{1 \le j \le k} f_j$, $\sup_j f_j$, $\inf_j f_j$, $\limsup_j f_j$ und $\liminf_j f_j$ zu $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$. Insbesondere ist $\lim_{j\to\infty} f_j \in \mathcal{L}_0(\Omega, \mathcal{A}, \bar{\mathbb{R}})$, falls der Limes in $\overline{\mathbb{R}}$ existiert.*

Beweis: Es sei $\alpha \in \mathbb{R}$. Dann ist nach Satz 4.3 $[f_j > \alpha] \in \mathcal{A}$, $j \in \mathbb{N}$, also

$$[\sup_j f_j > \alpha] = \bigcup_{j \ge 1} [f_j > \alpha] \in \mathcal{A},$$

also ist $\sup_j f_j$ messbar nach Satz 4.3. Ist $f_j \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$, so auch $-f_j$, also ist $\inf_j f_j = -\sup_j(-f_j)$ messbar.

Setze weiter $g_j = f_j$ für $1 \le j \le k$ und $g_j = f_k$ für $j > k$, so ist $\sup_j g_j = \max_{1 \le j \le k} f_j$ messbar, wie gerade gesehen, und analog ist auch $\min_{1 \le j \le k} f_j$ messbar. Nach Definition und mit $|f| = f^+ + f^-$ ist dann auch f^+, f^- und $|f|$ messbar. Weiter gilt

$$\limsup_j f_j = \inf_j \sup_{k \ge j} f_k, \quad \liminf_j f_j = \sup_j \inf_{k \ge j} f_k.$$

Also sind diese Abbildungen nach dem bereits Bewiesenen messbar. $\square$

4.6 Satz *Eine Funktion $f := (f_1, \dots, f_d) \colon (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}^d)$ ist genau dann $\mathcal{A}/\mathcal{B}^d$-messbar, wenn alle Koordinatenfunktionen $f_1, \dots, f_d \colon (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ $\mathcal{A}/\mathcal{B}$- messbar sind.*

Beweis: Es sei $\pi_j \colon \mathbb{R}^d \to \mathbb{R}$ definiert durch $\pi_j(x) := x_j$ für $x = (x_1, \dots, x_d)$ und $j \in \{1, \dots, d\}$ (Projektionsabbildungen). Jedes π_j ist stetig, also Borel-messbar. Ist f messbar, so auch $f_j := \pi_j \circ f$, $j = 1, \dots, d$ (nach Satz 3.5). Sind $f_1, \dots, f_d$ messbar und ist $[a, b[\in \mathcal{R}^d$, so ist

$$f^{-1}\big([a, b[\big) = \bigcap_{j=1}^{d} f_j^{-1}\big([a_j, b_j[\big) \in \mathcal{A}.$$

Da $\mathcal{R}^d$ die σ-Algebra $\mathcal{B}^d$ erzeugt, ist f mit Satz 3.3 messbar. $\square$

4.7 Satz *Sind f und g $\mathcal{A}$-messbare numerische Funktionen und $\alpha, \beta \in \overline{\mathbb{R}}$, so sind auch $\alpha f + \beta g$ und $f \cdot g$ $\mathcal{A}$-messbare numerische Funktionen.*

Beweis: Es seien $f, g \colon \Omega \to \mathbb{R}$. Dann ist $h \colon \Omega \to \mathbb{R}^2$, definiert durch $h(x) := (f(x), g(x))$, nach Satz 4.6 $\mathcal{A}/\mathcal{B}^2$-messbar. Seien weiter $s, p \colon \mathbb{R}^2 \to \mathbb{R}$ definiert durch $s(x_1, x_2) = x_1 + x_2$ und $p(x_1, x_2) = x_1 \cdot x_2$. Diese Abbildungen sind stetig, also $\mathcal{B}^2/\mathcal{B}$-messbar. Damit sind $f + g = s \circ h$ und $f \cdot g = p \circ h$ nach Satz 3.5 $\mathcal{A}/\mathcal{B}$-messbar.

Seien nun $f, g \colon \Omega \to \overline{\mathbb{R}}$ $\mathcal{A}$-messbare numerische Funktionen, so sind die Abbildungen $f_n, g_n \colon \Omega \to \mathbb{R}$ mit $f_n := \max\bigl(-n, \min(f, n)\bigr)$ und $g_n := \max\bigl(-n, \min(g, n)\bigr)$ nach Satz 4.5 $\mathcal{A}/\mathcal{B}$-messbar, also sind auch $f_n + g_n$ und $f_n \cdot g_n$, $n \in \mathbb{N}$, nach dem ersten Teil des Beweises $\mathcal{A}/\mathcal{B}$-messbar. Damit sind auch $f + g = \lim_{n\to\infty}(f_n + g_n)$ und $f \cdot g = \lim_{n\to\infty}(f_n \cdot g_n)$ $\mathcal{A}$-messbare numerische Funktionen (Satz 4.5). Konstante Funktionen α bzw. β sind messbar, also auch $\alpha f, \beta g$ und folglich $\alpha f + \beta g$. $\square$

4.8 Korollar *Sind $f, g \colon (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ $\mathcal{A}/\overline{\mathcal{B}}$-messbar, so sind $[f < g], [f \le g], [f = g]$ und $[f \ne g]$ in $\mathcal{A}$.*

Beweis: Es gelten $[f < g] = [g - f > 0]$, $[f \le g] = [g - f \ge 0]$, $[f = g] = [f - g \le 0] \cap [f - g \ge 0]$ und $[f \ne g] = [f - g < 0] \cup [f - g > 0]$. Also folgt alles aus der Messbarkeit von $f - g$ und $g - f$ und aus Satz 4.3. $\square$

4.9 Korollar *Eine numerische Funktion $f \colon (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ ist genau dann messbar, wenn ihr Positivteil f^+ und ihr Negativteil f^- messbar sind.*

Beweis: $\Rightarrow$ folgt aus Satz 4.5; $\Leftarrow$ folgt aus $f = f^+ - f^-$ und Satz 4.7. $\square$

4.10 Definition Es sei $(\Omega, \mathcal{A})$ eine Messraum. Eine Funktion $f \colon \Omega \to \mathbb{R}$ heißt $(\mathcal{A})$-*Elementarfunktion* oder $(\mathcal{A})$-*einfach*, wenn $f(\Omega)$ endlich ist und wenn sie $\mathcal{A}/\mathcal{B}$-messbar ist. Die Menge aller $(\mathcal{A})$-einfachen Funktionen bezeichnen wir mit $\mathcal{EF}(\Omega, \mathbb{R})$.

4.11 Bemerkungen

(a) Es seien $m \in \mathbb{N}$ und $(\alpha_j, A_j) \in \mathbb{R} \times \mathcal{A}$ für $j = 1, \ldots, m$. Dann ist

$$f := \sum_{j=1}^{m} \alpha_j 1_{A_j} \in \mathcal{EF}(\Omega, \mathbb{R}).$$

Gilt $\alpha_j \ne 0$, $j = 1, \ldots, m$, $\alpha_j \ne \alpha_k$ und $A_j \cap A_k = \emptyset$ für $j \ne k$, so heißt $\sum_{j=1}^{m} \alpha_j 1_{A_j}$ *Normalform* der einfachen Funktion f.

(b) Es gilt: Jede einfache Funktion besitzt eine eindeutig bestimmte Normalform und

$$\mathcal{EF}(\Omega, \mathbb{R}) = \left\{ \sum_{j=1}^{m} \alpha_j 1_{A_j}, m \in \mathbb{N}, \alpha_j \in \mathbb{R}\backslash\{0\}, A_j \in \mathcal{A}; A_j \cap A_k = \emptyset \text{ für } j \neq k \right\}.$$

(c) Mit $f, g \in \mathcal{EF}(\Omega, \mathbb{R})$ ist auch $|f|$, $g \cdot f$, $g \vee f$ und $g \wedge f$ in $\mathcal{EF}(\Omega, \mathbb{R})$.

Beweis zu (b): Es sei $f \in \mathcal{EF}(\Omega, \mathbb{R})$. Dann existiert ein $m \in \mathbb{N}$ und paarweise verschiedene $e_1, \ldots, e_m$ in $\mathbb{R}$ mit $f(\Omega) \setminus \{0\} = \{e_1, \ldots, e_m\}$. Sei $A_j := f^{-1}(e_j)$, dann ist $A_j \in \mathcal{A}$ und $A_j \cap A_k = \emptyset$ für $j \neq k$. Also ist $\sum_{j=1}^{m} e_j 1_{A_j}$ die eindeutig bestimmte Normalform. Der zweite Teil der Aussage folgt mit der ersten Aussage in Bemerkung (a). □

Es bezeichne $\overline{\mathbb{R}}^+ := [0, \infty]$. Wir können nun den Raum $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ charakterisieren:

4.12 Satz *Für $f \colon \Omega \to \overline{\mathbb{R}}^+$ sind äquivalent.*

(a) $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$

(b) *Es gibt eine wachsende Folge $(f_j)_j$ in $\mathcal{EF}(\Omega, \mathbb{R}^+)$ mit $f_j \to f$ für $j \to \infty$.*

Beweis: $(b) \Rightarrow (a)$: folgt sofort aus Satz 4.5.

$(a) \Rightarrow (b)$: Es sei $j \in \mathbb{N}$ und

$$A_{j,k} := \begin{cases} [k2^{-j} \leq f < (k+1)2^{-j}] & , \ k = 0, \ldots, j2^j - 1 \\ [f \geq j] & , \ k = j2^j. \end{cases}$$

Die Mengen $A_{j,k}$ sind für $k = 0, \ldots, j2^j$ disjunkt und liegen in $\mathcal{A}$ gemäß Satz 4.3. Dann gehört

$$f_j := \sum_{k=0}^{j2^j} k2^{-j} \ 1_{A_{j,k}}, \ j \in \mathbb{N},$$

zu $\mathcal{EF}(\Omega, \mathbb{R}^+)$. f_{j+1} kann auf $A_{j,k}$ nur die Werte $2^{-j-1}(2k)$ und $2^{-j-1}(2k+1)$ für $k = 0, \ldots, j2^j - 1$ annehmen, denn:

$$A_{j,k} = \left[\frac{2k}{2^{j+1}} \leq f < \frac{2k+1}{2^{j+1}}\right] \cup \left[\frac{2k+1}{2^{j+1}} \leq f < \frac{2(k+1)}{2^{j+1}}\right] = A_{j+1,2k} \cup A_{j+1,2k+1}.$$

Dann ist

$$f_{j+1}|_{A_{j+1,2k}} = (2k)\, 2^{-j-1} = f_j|_{A_{j,k}}$$

und

$$f_{j+1}|_{A_{j+1,2k+1}} = (2k+1)\, 2^{-j-1} > f_j|_{A_{j,k}} \ .$$

Auf $A_{j,j\,2^j}$ kann f_{j+1} nur Werte $\geq j$ annehmen:

$$[f \geq j] = [f \geq j+1] \cup [j \leq f < j+1].$$

Also folgt

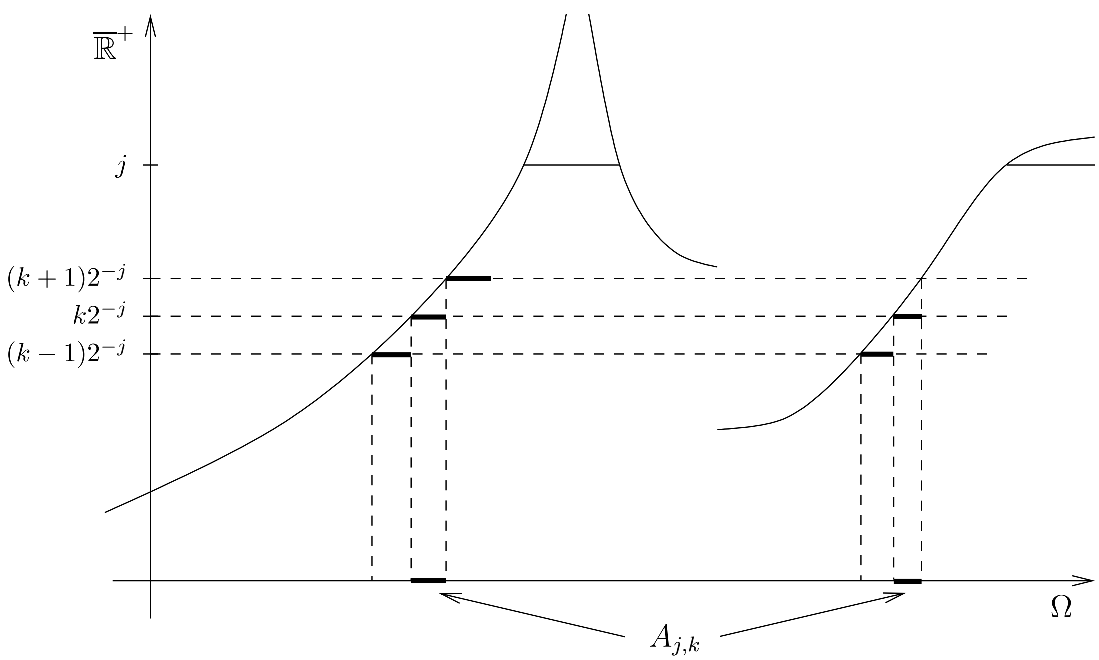
$$f_{j+1}|_{[f \geq j+1]} = j+1 > f_j|_{[f \geq j+1]} = j$$

und

$$f_{j+1}|_{[j \leq f < j+1]} \geq j = f_j|_{[j \leq f < j+1]} \ .$$

Daher ist $0 \leq f_j \leq f_{j+1}$ für $j \in \mathbb{N}$.

Es sei nun $x \in \Omega$. Ist $f(x) = +\infty$, so gilt $f_j(x) = j$, $j \in \mathbb{N}$, und somit $\lim_{j\to\infty} f_j(x) = f(x)$. Ist $f(x) < \infty$, so gilt für $j > f(x)$, dass x in einer Menge $A_{j,k}$, $k = 0, \ldots, j\,2^j - 1$ liegt. Es gilt $f_j(x) \leq f(x) < f_j(x) + 2^{-j}$, also auch hier $\lim_{j\to\infty} f_j(x) = f(x)$. Somit ist die punktweise Konvergenz gezeigt.



□

4.13 Korollar

(a) *Zu jedem $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ gibt es eine Folge $(f_j)_j$ in $\mathcal{EF}(\Omega, \mathbb{R})$ mit $f_j \to f$.*

(b) *Es sei $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R}^+)$ eine beschränkte Abbildung. Dann gibt es eine wachsende Folge $(f_j)_j$ in $\mathcal{EF}(\Omega, \mathbb{R}^+)$, die gleichmäßig gegen f konvergiert.*

Beweis: Zu (a): Es gilt $f = f^+ - f^-$. Somit folgt die Behauptung aus Satz 4.12 und der Tatsache, das

$$\mathcal{EF}(\Omega, \mathbb{R}) \subset \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$$

im Sinne von Untervektorräumen.

Zu (b): Wenn $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R}^+)$ eine beschränkte Abbildung ist, so folgt für die im Beweis von Satz 4.12 konstruierte Folge $(f_j)_j$:

$$f_j(x) \leq f(x) < f_j(x) + 2^{-j}$$

für $j > \|f\|_\infty$, also konvergiert $(f_j)_j$ gegen f gleichmäßig. $\square$

KAPITEL 5

Integrierbare Funktionen

Entlang der Diskussion messbarer numerischer Funktionen, definiert auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$, ergibt sich die Konstruktion des Integrals von numerischen Funktionen bezüglich des Maßes μ in drei Schritten. Für $f = 1_A$, $A \in \mathcal{A}$, soll $\int f d\mu := \mu(A)$ sein. Für einfache Funktionen $f = \sum_{j=1}^m \alpha_j \, 1_{A_j}$ wird dann $\sum_{j=1}^m \alpha_j \, \mu(A_j)$ der Kandidat für das Integral sein. Für nicht–negative messbare numerische Funktionen wollen wir dann die Approximierbarkeit durch eine wachsende Folge von einfachen Funktionen nutzen. Schliesslich wird für eine beliebige messbare numerische Funktion ihre Zerlegung in den Positiv– und Negativteil verwendet.

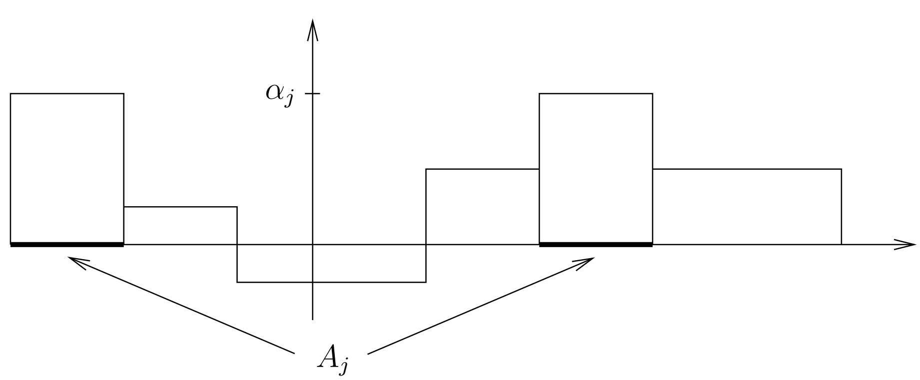
Wir vereinbaren im folgenden einfache Funktionen stets durch ihre Normalform darzustellen, wenn es nicht ausdrücklich anders gesagt wird. Ferner setzen wir $\infty \cdot 0 := -\infty \cdot 0 := 0$. Diese Vereinbarung ist in der Integrationstheorie üblich und nützlich, um zum Beispiel einfache Funktionen über ihren ganzen Definitionsbereich integrieren zu können. Es ist eine Multiplikation der Elemente $+\infty$ und $-\infty$ mit dem Nullvektor 0 in $\mathbb{R}$ (oder allgemeiner dem Nullvektor 0 eines Vektorraums), und keine weitere Rechenregel in $\overline{\mathbb{R}}$. Wir erinnern daran, dass $\overline{\mathbb{R}}$ kein Körper ist und in $\overline{\mathbb{R}}$ Ausdrücke wie $\infty - \infty$, $0 \cdot (\pm\infty)$, $\frac{\pm\infty}{\infty}$, $\frac{\pm\infty}{-\infty}$, $\frac{0}{0}$ und $\frac{\pm\infty}{0}$ *nicht definiert* sind.

Ferner sei $(\Omega, \mathcal{A}, \mu)$ stets ein Maßraum.

5.1 Definition Für $f = \sum_{j=1}^m \alpha_j 1_{A_j} \in \mathcal{EF}(\Omega, \mathbb{R})$ (Normalform) heißt

$$\int_\Omega f d\mu := \int f d\mu := \sum_{j=1}^m \alpha_j \mu(A_j)$$

Integral von f (über Ω) bezüglich des Maßes μ. Weiter heißt $\int_A f d\mu := \int_\Omega 1_A f d\mu$ für $A \in \mathcal{A}$ *Integral* von f über A bezüglich des Maßes μ.



5.2 Bemerkungen

(a) Nach der Bemerkung 4.11 (b) ist $\int f\,d\mu$ wohldefiniert. Nach Bemerkung 4.11 (c) ist auch $\int_A f\,d\mu$ für $A \in \mathcal{A}$ wohldefiniert.

(b) Ist $f = \sum_{k=1}^{n} \beta_k\, 1_{B_k}$ mit $\beta_1, \ldots, \beta_n \in \mathbb{R} \setminus \{0\}$ und $B_1, \ldots, B_n \in \mathcal{A}$ mit $B_j \cap B_k = \emptyset$ für $j \neq k$ (nicht notwendig die Normalform!) eine $\mathcal{A}$-einfache Funktion, so gilt

$$\int_\Omega f\,d\mu = \sum_{k=1}^{n} \beta_k\, \mu(B_k).$$

Beweis zu (b): Es sei $\sum_{j=1}^{m} \alpha_j\, 1_{A_j}$ die Normalform von f. Sei $A_{m+1} = \bigcap_{j=1}^{m} A_j^c$, $B_{n+1} = \bigcap_{k=1}^{n} B_k^c$ und $\alpha_{m+1} = 0$, $\beta_{n+1} = 0$. Dann gilt $\Omega = \bigcup_{j=1}^{m+1} A_j = \bigcup_{k=1}^{n+1} B_k$ und $A_j = \bigcup_{k=1}^{n+1} (A_j \cap B_k)$ und $B_k = \bigcup_{j=1}^{m+1} (B_k \cap A_j)$ für $j = 1, \ldots, m$ und $k = 1, \ldots, n$. Die Mengen $A_j \cap B_k$ sind paarweise disjunkt. Daher folgt

$$\mu(A_j) = \sum_{k=1}^{n+1} \mu(A_j \cap B_k)\ , \quad \mu(B_k) = \sum_{j=1}^{m+1} \mu(B_k \cap A_j).$$

Ist $A_j \cap B_k \neq \emptyset$, so ist $\alpha_j = \beta_k$, also

$$\begin{aligned} \int_\Omega f\,d\mu &= \sum_{j=1}^{m} \alpha_j\, \mu(A_j) = \sum_{j=1}^{m+1} \alpha_j \sum_{k=1}^{n+1} \mu(A_j \cap B_k) \\ &= \sum_{k=1}^{n+1} \beta_k \sum_{j=1}^{m+1} \mu(A_j \cap B_k) = \sum_{k=1}^{n} \beta_k\, \mu(B_k)\ . \end{aligned}$$

□

5.3 Satz *Das Integral* $\int_\Omega \cdot\, d\mu \colon \mathcal{EF}(\Omega, \mathbb{R}) \to \mathbb{R}$ *ist linear. Für* $A, B \in \mathcal{A}$ *und* $A \cap B = \emptyset$ *gilt*

$$\int_{A \cup B} f\,d\mu = \int_A f\,d\mu + \int_B f\,d\mu, \quad f \in \mathcal{EF}(\Omega, \mathbb{R}).$$

Weiter gilt

$$\left| \int_A f\,d\mu \right| \leq \int_A |f|\,d\mu \leq \|f\|_\infty\, \mu(A), \quad A \in \mathcal{A}, \ f \in \mathcal{EF}(\Omega, \mathbb{R})$$

und $\int_A f\,d\mu \leq \int_A g\,d\mu$ *für* $f, g \in \mathcal{EF}(\Omega, \mathbb{R})$ *mit* $f \leq g$ *und* $A \in \mathcal{A}$.

Beweis: Die Gleichheit $\int_\Omega \alpha f\,d\mu = \alpha \int_\Omega f\,d\mu$ für $\alpha \in \mathbb{R}$ ist sofort klar. In der Notation wie im Beweis der Bemerkung 5.2 (b) gilt

$$1_{A_j} = \sum_{k=1}^{n+1} 1_{A_j \cap B_k} \text{ und } 1_{B_k} = \sum_{j=1}^{m+1} 1_{A_j \cap B_k}.$$

Wenn nun $f = \sum_{j=1}^{m} \alpha_j 1_{A_j}$ und $g = \sum_{k=1}^{n} \beta_k 1_{B_k}$, so folgt

$$f + g = \sum_{j=1}^{m+1} \sum_{k=1}^{n+1} (\alpha_j + \beta_k) 1_{A_j \cap B_k}.$$

Dies ist im Allgemeinen keine Normalform. Doch mit Bemerkung 5.2 (b) folgt die behauptete Linearität. Da weiter $1_{A \cup B}\, f = 1_A\, f + 1_B\, f$ gilt, folgt somit die zweite Aussage. Die dritte Aussage folgt, da mit f auch $|f|$ einfach ist und die Dreiecksungleichung gilt. Weiter ist $\int_A h\, d\mu \geq 0$ für $h \in \mathcal{EF}(\Omega, \mathbb{R}^+)$, also folgt mit der Linearität auch die letzte Behauptung. □

Bisher brauchten wir nur die endliche Additivität von μ. Die Resultate 5.2 und 5.3 gelten auch für Inhalte auf Ringen und Algebren. Wir wollen nun das Integral für nicht-negative $\mathcal{A}$-messbare numerische Funktionen bestimmen. Nach Satz 4.12 existiert zu $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ eine Folge $(f_j)_j$ in $\mathcal{EF}(\Omega, \mathbb{R}^+)$ mit $f_j \uparrow f$. Es bietet sich daher die Definition $\int_\Omega f\, d\mu := \lim_{j\to\infty} \int_\Omega f_j\, d\mu$ an. Tatsächlich ist diese Definition sinnvoll, da sie unabhängig von der Wahl der approximierenden Folge $(f_j)_j$ ist:

5.4 Satz *Für jede wachsende Folge $(f_j)_j$ in $\mathcal{EF}(\Omega, \mathbb{R}^+)$ und jedes $f \in \mathcal{EF}(\Omega, \mathbb{R}^+)$ mit $f \leq \sup_j f_j$ gilt: $\int_\Omega f\, d\mu \leq \sup_j \int_\Omega f_j\, d\mu$.*

Beweis: Es seien $f = \sum_{i=1}^{m} \alpha_i 1_{A_i}$, $\alpha_1, \ldots, \alpha_m \geq 0$ und $A_1, \ldots, A_m \in \mathcal{A}$ disjunkt. Zu $\beta > 1$ und $j \in \mathbb{N}$ sei $B_j := [\beta\, f_j \geq f]$. Nach Korollar 4.8 liegen diese Mengen in der σ-Algebra $\mathcal{A}$. Ist $x \in \Omega$ mit $f(x) = 0$, so ist $x \in B_j$ für alle $j \in \mathbb{N}$. Ist $f(x) > 0$, so ist $\lim_{k\to\infty} \beta\, f_k(x) > f(x)$. Also ist $x \in B_j$ für hinreichend große $j \in \mathbb{N}$. Somit folgt $B_j \uparrow \Omega$. Nach Definition von B_j ist $\beta\, f_j \geq f\, 1_{B_j}$. Es folgt mit Satz 1.19

$$\begin{aligned} \int_\Omega f\, d\mu &= \sum_{i=1}^{m} \alpha_i\, \mu(A_i) = \lim_{j\to\infty} \sum_{i=1}^{m} \alpha_i\, \mu(A_i \cap B_j) \\ &= \lim_{j\to\infty} \int_\Omega f\, 1_{B_j}\, d\mu \leq \lim_{j\to\infty} \int_\Omega \beta\, f_j\, d\mu = \beta\, \lim_{j\to\infty} \int_\Omega f_j\, d\mu. \end{aligned}$$

Für $\beta \downarrow 1$ folgt die Behauptung. □

5.5 Korollar *Für zwei wachsende Folgen $(f_j)_j$, $(g_j)_j$ in $\mathcal{EF}(\Omega, \mathbb{R}^+)$ gilt:*
Ist $\sup_j f_j = \sup_j g_j$, so folgt

$$\sup_j \int_\Omega f_j\, d\mu = \sup_j \int_\Omega g_j\, d\mu.$$

Beweis: Für $k \in \mathbb{N}$ ist $g_k \le \sup_j f_j$, also mit Satz 5.4

$$\int_\Omega g_k \, d\mu \le \sup_j \int_\Omega f_j \, d\mu, \text{ also } \sup_k \int_\Omega g_k \, d\mu \le \sup_j \int_\Omega f_j \, d\mu.$$

Die Symmetrie dieses Arguments liefert die Behauptung. □

5.6 Definition Es seien $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ und $(f_j)_j$ eine Folge aus $\mathcal{EF}(\Omega, \mathbb{R}^+)$ mit $f_j \uparrow f$. Dann heißt der von der Auswahl der Folge $(f_j)_j$ unabhängige Grenzwert

$$\int_\Omega f \, d\mu := \lim_{j \to \infty} \int_\Omega f_j \, d\mu = \sup_j \int_\Omega f_j \, d\mu$$

das *Integral* von f (über Ω) bezüglich des Maßes μ, kurz das μ-Integral von f. Wir verwenden im Weiteren auch die Symbole

$$\int f \, d\mu, \quad \int_\Omega f(x) \, d\mu(x), \quad \int_\Omega f(x) \, \mu(dx).$$

Wir haben in Satz 5.3 gesehen, dass das Integral auf dem Raum der einfachen Funktionen linear und monoton ist. Wir zeigen nun, dass diese Eigenschaften bei der Ausdehnung des Integrals auf den Raum $\mathcal{L}_0(\Omega, \mu, \overline{\mathbb{R}}^+)$ erhalten bleiben:

5.7 Satz

(a) *Für* $f, g \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ *und* $\alpha, \beta \in \mathbb{R}^+$ *gilt*

$$\int_\Omega (\alpha \, f + \beta \, g) \, d\mu = \alpha \int_\Omega f \, d\mu + \beta \int_\Omega g \, d\mu.$$

(b) *Für* $f, g \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ *mit* $f \le g$ *gilt*

$$\int_\Omega f \, d\mu \le \int_\Omega g \, d\mu.$$

Beweis: zu (a): Zu f und g existieren Folgen $(f_j)_j$ und $(g_j)_j$ in $\mathcal{EF}(\Omega, \mathbb{R}^+)$ mit $f_j \uparrow f$ und $g_j \uparrow g$, also $f_j + g_j \uparrow f + g$. Somit

$$\int (f+g) \, d\mu = \lim_{j \to \infty} \int (f_j+g_j) \, d\mu = \lim_{j \to \infty} \int f_j \, d\mu + \lim_{j \to \infty} \int g_j \, d\mu = \int f \, d\mu + \int g \, d\mu$$

nach Satz 5.3. Für $\alpha \in \mathbb{R}^+$ folgt analog $\int \alpha \, f \, d\mu = \alpha \int f \, d\mu$.

zu (b): Mit $g = f + (g - f)$ und $g - f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ ergibt (i):

$$\int g \, d\mu = \int f \, d\mu + \int (g - f) \, d\mu \ge \int f \, d\mu.$$

□

5.8 Bemerkung Es gilt

$$\mathcal{E}\mathcal{F}(\Omega, \mathbb{R}^+) \subset \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+).$$

Denn für $f \in \mathcal{E}\mathcal{F}(\Omega, \mathbb{R}^+)$ gilt $\lim_{j\to\infty} f_j = f$, wenn wir $f_j = f$, $j \in \mathbb{N}$, setzen. Somit ist die Definition 5.6 mit Definition 5.1 verträglich.

5.9 Korollar *Für $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ gilt*

$$\int_\Omega f\, d\mu = \sup\left\{ \int_\Omega g\, d\mu,\ g \in \mathcal{E}\mathcal{F}(\Omega, \mathbb{R}^+),\ g \le f \right\}$$

(dies hätte die anfängliche Definition sein können).

Beweis: Für alle $g \in \mathcal{E}\mathcal{F}(\Omega, \mathbb{R}^+)$ mit $g \le f$ ist $\int_\Omega f\, d\mu \ge \int_\Omega g\, d\mu$ (Monotonie, Satz 5.3), also

$$\int_\Omega f\, d\mu \ge \sup\left\{ \int_\Omega g\, d\mu,\ g \in \mathcal{E}\mathcal{F}(\Omega, \mathbb{R}^+),\ g \le f \right\}.$$

Die andere Richtung folgt sofort aus der Definition 5.6. □

Satz 5.4 gilt auch in $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ und geht auf B. Levi (1875 - 1961) zurück:

5.10 Satz (über die monotone Konvergenz von Levi) *Es sei $(f_j)_j$ eine wachsende Folge in $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$. Dann gilt*

$$\int_\Omega \big(\lim_{j\to\infty} f_j\big)\, d\mu = \lim_{j\to\infty} \int_\Omega f_j\, d\mu \quad in \quad \overline{\mathbb{R}}^+.$$

Beweis: Es sei $f := \lim_{j\to\infty} f_j$, dann ist nach Satz 4.5 $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ und es ist $f_j \le f$, $j \in \mathbb{N}$. Also folgt aus Satz 5.7 (Monotonie) $\int f_j\, d\mu \le \int f\, d\mu$, $j \in \mathbb{N}$, also $\lim_{j\to\infty} \int f_j\, d\mu \le \int f\, d\mu$.

Es sei nun $g \in \mathcal{E}\mathcal{F}(\Omega, \mathbb{R}^+)$ mit $g \le f$. Sei $\beta > 1$ und $B_j := [\beta f_j \ge g]$, $j \in \mathbb{N}$. Dann ist $(B_j)_j$ eine wachsende Folge in $\mathcal{A}$ mit $\bigcup_j B_j = \Omega$ und $\beta f_j \ge g\, 1_{B_j}$ (siehe Beweis zu Satz 5.4). Es ist $g\, 1_{B_j} \uparrow g$, also folgt mit Satz 5.3

$$\int g\, d\mu = \lim_{j\to\infty} \int g\, 1_{B_j}\, d\mu \le \beta \lim_{j\to\infty} \int f_j\, d\mu\ .$$

Da $\beta > 1$ beliebig war, folgt weiter

$$\int g\, d\mu \le \lim_{j\to\infty} \int f_j\, d\mu$$

für jede einfache Funktion g mit $g \leq f$. Mit Korollar 5.9 folgt

$$\int f \, d\mu \leq \lim_{j \to \infty} \int f_j \, d\mu \, .$$

$\square$

5.11 Korollar *Es sei* $(f_j)_j$ *eine Folge in* $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$. *Dann gilt*

$$\sum_{j=1}^{\infty} \int f_j \, d\mu = \int \big(\sum_{j=1}^{\infty} f_j \big) \, d\mu \quad \textit{in} \quad \overline{\mathbb{R}}^+.$$

Beweis: Man wende Satz 5.10 für $(f_1 + \cdots + f_j)_{j \in \mathbb{N}}$ an und verwende Satz 5.7 (Additivität des Integrals). $\square$

5.12 Bemerkung Die Aussage von Satz 5.10 über die monotone Konvergenz ist für nicht-wachsende Folgen im allgemeinen falsch. Sei dazu $f_j = \frac{1}{j} 1_{[0,j]}$, $j \in \mathbb{N}$. Dann ist $(f_j)_j$ eine Folge in $\mathcal{EF}(\mathbb{R}, \mathbb{R}^+)$. In Bezug auf den Maßraum $(\mathbb{R}, \mathcal{B}^1, \lambda^1)$ gilt: $\int f_j \, d\lambda^1 = 1$, $j \in \mathbb{N}$. Aber $(f_j)_j$ konvergiert gleichmäßig auf ganz $\mathbb{R}$ gegen 0.

5.13 Beispiele

(a) Es seien $(\Omega, \mathcal{A})$ ein Messraum und δ_ω das Dirac–Maß (siehe Beispiel 1.13 (a)). Dann ist $\int f \, d\delta_\omega = f(\omega)$ für jedes $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$. Nach Definition 5.6 genügt es, $f \in \mathcal{EF}(\Omega, \mathbb{R}^+)$ zu betrachten. Sei also $f = \sum_{i=1}^m \alpha_i \, 1_{A_i}$ in Normaldarstellung gegeben. Dann liegt ω in genau einer Menge A_j, also folgt

$$\int f \, d\delta_\omega = \sum_{i=1}^{m} \alpha_i \, \delta_\omega(A_i) = \alpha_j = f(\omega).$$

(b) Es seien $\Omega = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ und ϱ das Zählmaß (siehe Beispiel 1.13 (b)). Dann ist $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ gleich der Menge aller numerischen Funktionen $f \geq 0$ auf Ω. Weiter ist

$$\int f \, d\varrho = \sum_{n=1}^{\infty} f(n).$$

Setze dazu $f_n := f(n) \, 1_{\{n\}}$, $n \in \mathbb{N}$. Dann ist $f_n \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$, sogar in $\mathcal{EF}(\Omega, \mathbb{R}^+)$, und mit $f = \sum_{n \geq 1} f_n$ folgt aus Korollar 5.11 $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ und

$$\int f \, d\varrho = \sum_{n=1}^{\infty} \int f_n \, d\varrho = \sum_{n=1}^{\infty} f(n) \, \varrho(\{n\}) = \sum_{n=1}^{\infty} f(n).$$

Wählen wir $f_n\colon \mathbb{N} \to [0,\infty]$, $n \in \mathbb{N}$, mit $f_n(k) = a_{nk} \in \mathbb{R}$, $n,k \in \mathbb{N}$. Dann liefert Korollar 5.11 für alle $a_{nk} \in [0,\infty]$ $(n,k \in \mathbb{N})$

$$\sum_{k=1}^{\infty}\Big(\sum_{n=1}^{\infty} a_{nk}\Big) = \sum_{n=1}^{\infty}\Big(\sum_{k=1}^{\infty} a_{nk}\Big).$$

Dies ist für nicht-negative Doppelreihen sogar eine Erweiterung einer üblichen Formulierung, da nicht $\sup_l \sum_{n,k=0}^{l} |a_{nk}| < \infty$ gefordert wird.

Wir beweisen nun eine Verallgemeinerung des Satzes über die monotone Konvergenz für beliebige nicht notwenigerweise wachsende Folgen in $\mathcal{L}_0(\Omega,\mathcal{A},\overline{\mathbb{R}}^+)$:

5.14 Lemma (von Fatou) *Für jede Folge $(f_j)_j$ in $\mathcal{L}_0(\Omega,\mathcal{A},\overline{\mathbb{R}}^+)$ gilt*

$$\int \Big(\liminf_j f_j\Big)\, d\mu \le \liminf_j \int f_j\, d\mu \quad \textit{in} \quad \overline{\mathbb{R}}^+.$$

Beweis: Wir setzen $g_j = \inf_{k\ge j} f_k$. Nach Satz 4.5 ist $g_j \in \mathcal{L}_0(\Omega,\mu,\overline{\mathbb{R}}^+)$ für jedes $j \in \mathbb{N}$. Die Folge $(g_j)_j$ konvergiert wachsend gegen $\liminf_j f_j$. Also liefert Satz 5.10

$$\lim_{j\to\infty} \int g_j\, d\mu = \int \Big(\liminf_j f_j\Big)\, d\mu.$$

Weiter ist $g_j \le f_k$, also $\int g_j\, d\mu \le \int f_k\, d\mu$ für $k \ge j$. Also $\int g_j\, d\mu \le \inf_{k\ge j} \int f_k\, d\mu$. Im Limes $j \to \infty$ folgt nun die Behauptung. □

Nun dehnen wir den Integralbegriff auf geeignete messbare Funktionen aus:

5.15 Definition Man nennt $f \in \mathcal{L}_0(\Omega,\mathcal{A},\overline{\mathbb{R}})$ *μ-integrierbar*, wenn $\int_\Omega f^+\, d\mu < \infty$ und $\int_\Omega f^-\, d\mu < \infty$. Äquivalent dazu ist die Bedingung $\int_\Omega |f|\, d\mu < \infty$. In diesem Fall heißt

$$\int_\Omega f\, d\mu := \int_\Omega f^+\, d\mu - \int_\Omega f^-\, d\mu$$

das (μ-) *Integral* von f (über Ω). Ist f μ-integrierbar und $A \in \mathcal{A}$, so setzt man $\int_A f\, d\mu := \int 1_A f\, d\mu$ und nennt dies das μ-Integral von f über A. Wenn keine Verwechslungen möglich sind, lassen wir das Maß μ in der Bezeichnung weg und sprechen vom Integral von f und der Integrierbarkeit von f.

5.16 Bemerkungen

(a) Für $f \in \mathcal{L}_0(\Omega,\mathcal{A},\overline{\mathbb{R}}^+)$ ist $f^- = 0$, und somit ist Definition 5.15 mit Definition 5.6 verträglich.

(b) Für $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ gilt: f ist μ-integrierbar genau dann, wenn $\int f \, d\mu < \infty$.

(c) $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ heißt *quasi-integrierbar*, falls $\int f^+ \, d\mu < \infty$ oder $\int f^- \, d\mu < \infty$. Wieder setzt man $\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu \in \overline{\mathbb{R}}$.

5.17 Satz *Für die μ-Integrierbarkeit von $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ ist jede der folgenden Bedingungen notwendig und hinreichend:*

(a) *f^+ und f^- sind μ-integrierbar.*

(b) *Es gibt μ-integrierbare Funktionen $u \geq 0$, $v \geq 0$ mit $f = u - v$.*

(c) *Es gibt eine μ-integrierbare Funktion g mit $|f| \leq g$.*

(d) *$|f|$ ist μ-integrierbar.*

Beweis: Wir zeigen, dass $(a) - (d)$ äquivalent sind, denn die μ-Integrierbarkeit von f ist äquivalent zu (a).

$(a) \Rightarrow (b)$: Man setzt $u := f^+$, $v = f^-$ und verwendet $f = f^+ - f^-$.

$(b) \Rightarrow (c)$: Mit u, v ist auch $u + v$ integrierbar (nach Satz 5.7). Es gilt

$$f = u - v \leq u \leq u + v$$

und

$$-f = v - u \leq v \leq u + v,$$

also setzt man $g := u + v$.

$(c) \Rightarrow (d)$: Nach Satz 5.7 gilt $\int |f| \, d\mu \leq \int g \, d\mu < +\infty$.

$(d) \Rightarrow (a)$: Es gilt $f^+ \leq |f|$ und $f^- \leq |f|$ und mit Satz 5.7 folgt die Behauptung. $\square$

5.18 Bemerkung Aus (b) in Satz 5.17 folgt $\int f \, d\mu = \int u \, d\mu - \int v \, d\mu$, denn $f = u - v = f^+ - f^-$, also $u + f^- = v + f^+$. Mit Satz 5.7 folgt $\int u \, d\mu + \int f^- \, d\mu = \int v \, d\mu + \int f^+ \, d\mu$. $\square$

5.19 Korollar *Ist $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ μ-integrierbar, dann ist $\mu[|f| = \infty] = 0$. Sind $f, g \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ μ-integrierbar, so sind auch $\max(f, g)$ und $\min(f, g)$ μ-integrierbar.*

Beweis: Es gilt $A := [|f| = \infty] \in \mathcal{A}$ und $\infty \cdot 1_A \leq |f|$, also mit Satz 5.7

$$\infty \cdot \mu(A) = \int_\Omega \infty \cdot 1_A \, d\mu \leq \int_\Omega |f| \, d\mu < \infty.$$

Somit folgt $\mu(A) = 0$. Es sind $\max(f, g)$ und $\min(f, g)$ in $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ (siehe Satz 4.5). Beide Funktionen werden durch $|f| + |g|$ majorisiert, und $|f| + |g|$ ist eine μ-integrierbare Funktion. Also folgt die Behauptung mit Satz 5.17. □

5.20 Satz (Linearität und Monotonie des Integrals)

(a) *Sind $f, g \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ μ-integrierbar und $\alpha \in \mathbb{R}$, so ist $\alpha f + g$ μ-integrierbar und $\int (\alpha f + g)\, d\mu = \alpha \int f\, d\mu + \int g\, d\mu$ (hierfür muß $\alpha f + g$ wohldefiniert sein).*

(b) *Sind $f, g \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ μ-integrierbar, so gilt*

$$f \leq g \Rightarrow \int f d\mu \leq \int g d\mu \quad \text{und} \quad \Big| \int f d\mu \Big| \leq \int |f| d\mu.$$

Beweis: zu (a): Die Funktion αf ist μ-integrierbar, denn sie ist messbar und $|\alpha f| = |\alpha||f|$ ist nach Satz 5.7 μ-integrierbar. Es gilt $(\alpha f)^+ = \alpha f^+$ und $(\alpha f)^- = \alpha f^-$ für $\alpha \geq 0$ und $(\alpha f)^+ = |\alpha| f^-$ und $(\alpha f)^- = |\alpha| f^+$ für $\alpha \leq 0$. Damit folgt

$$\int \alpha f\, d\mu = \alpha \int f\, d\mu.$$

Weiter ist $f = f^+ - f^-$ und $g = g^+ - g^-$, also $f + g = f^+ + g^+ - (f^- + g^-)$. Mit Satz 5.7 sind $u := f^+ + g^+$ und $v := f^- + g^-$ μ-integrierbar. Also ist $f + g$ μ-integrierbar nach Satz 5.17 und die $f + g$ betreffende Behauptung folgt aus der Gleichheit $f + g = u - v$.

zu (b): Aus $f \leq g$ folgt $f^+ \leq g^+$ und $f^- \geq g^-$, also $\int f d\mu \leq \int g d\mu$ mit Satz 5.7. Da $f \leq |f|$ und $-f \leq |f|$, folgt $|\int f d\mu| \leq \int |f| d\mu$ aus dem gerade Bewiesenen, indem man speziell $g := |f|$ wählt. □

5.21 Definition Es bezeichne $\mathcal{L}^1 := \mathcal{L}^1(\mu) := \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ die Menge der reellwertigen μ-integrierbaren Funktionen.

5.22 Bemerkung $\mathcal{L}^1$ ist ein $\mathbb{R}$-Vektorraum und $f \mapsto \int f\, d\mu$ ist eine *monotone Linearform*.

5.23 Bemerkung Ist f μ-integrierbar, so auch $1_A \cdot f$ mit $A \in \mathcal{A}$, denn $|1_A \cdot f| \leq |f|$. Man verwende Satz 5.17. Damit ist $\int_A f d\mu := \int_\Omega 1_A\, f\, d\mu$ in Definition 5.15 ordentlich definiert. Die Abbildung $f \mapsto \int_A f d\mu$ ist eine Linearform auf $\mathcal{L}^1(\mu)$, $A \in \mathcal{A}$. Man kann Integrale über Mengen aus $\mathcal{A}$ auch gewinnen, indem man die Restriktion $\mu_A := \mu|_{A \cap \mathcal{A}}$ von μ auf die Spur σ-Algebra $A \cap \mathcal{A}$ betrachtet: Ist f' die Restriktion von f auf $A \in \mathcal{A}$, dann ist f' μ_A-integrierbar und

$$\int f'\, d\mu_A = \int_A f\, d\mu$$

(einfache Übung in drei Schritten).

5.24 Beispiele

(a) Siehe Beispiel 5.13 (a): $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ ist genau dann δ_ω-integrierbar, wenn $f(\omega) \in \mathbb{R}$.

(b) Siehe Beispiel 5.13 (b): $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ ist ϱ-integrierbar (ϱ bezeichne das Zählmaß), wenn

$$\sum_{n \geq 1} |f(n)| < \infty.$$

Es sei nun $(\Omega', \mathcal{A}')$ ein Messraum und $T\colon \Omega \to \Omega'$ eine $\mathcal{A}/\mathcal{A}'$-messbare Abbildung und $\mu' = T(\mu)$ das Bildmaß (siehe Definition 3.8). Dann gilt:

5.25 Satz

(a) *Für* $f' \in \mathcal{L}_0(\Omega', \mathcal{A}', \overline{\mathbb{R}}^+)$ *gilt*

$$\int_{\Omega'} f' \, dT(\mu) = \int_\Omega f' \circ T \, d\mu. \tag{5.1}$$

(b) *Es sei* $f' \in \mathcal{L}_0(\Omega', \mathcal{A}', \overline{\mathbb{R}})$. *Dann ist* f' *genau dann* $T(\mu)$*-integrierbar, wenn* $f' \circ T$ μ*-integrierbar ist und es gilt* (5.1).

(c) *Zu* $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ *wird durch* $\nu(A) := \int_A f \, d\mu$ *ein Maß auf* $\mathcal{A}$ *definiert (Maß mit der Dichte* f *bezüglich* μ*). Wir schreiben* $\nu = f\mu$*. Es gilt für jedes* $\varphi \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$

$$\int_\Omega \varphi \, d\nu = \int_\Omega \varphi f \, d\mu. \tag{5.2}$$

Eine $\mathcal{A}$*-messbare numerische Funktion* $\varphi : \Omega \to \overline{\mathbb{R}}$ *ist genau dann* ν*-integrierbar, wenn* φf μ*-integrierbar ist, und in diesem Fall gilt* (5.2).

Beweis: zu (a): Die Abbildung $f' \circ T$ ist $\mathcal{A}$-messbar, also ist $\int f' \circ T \, d\mu$ definiert. Es sei zunächst $f' = \sum_{i=1}^n \alpha_i 1_{A'_i}$ mit $A'_i \in \mathcal{A}'$ und $\alpha_i \in \mathbb{R}^+$, dann ist $f' \circ T = \sum_{i=1}^n \alpha_i 1_{A_i}$ mit $A_i = T^{-1}(A'_i)$. Mit $T(\mu)(A'_i) = \mu(A_i)$ für $i = 1, \dots, n$ folgt somit (5.1). Ist nun $f' \in \mathcal{L}_0(\Omega', \mathcal{A}', \overline{\mathbb{R}}^+)$, so existiert eine Folge $(f'_n)_n$ mit $f'_n \in \mathcal{EF}(\Omega', \mathbb{R}^+)$, $n \in \mathbb{N}$, und $f'_n \uparrow f'$. Dann ist $(f'_n \circ T)_n$ eine Folge in $\mathcal{EF}(\Omega, \mathbb{R}^+)$ mit $f'_n \circ T \uparrow f' \circ T$. Die Definition des Integrals liefert dann unmittelbar (5.1).

zu (b): Nach Teil (a) ist $\int (f')^+ \, dT(\mu) = \int (f')^+ \circ T \, d\mu$ und $\int (f')^- \, dT(\mu) = \int (f')^- \circ T d\mu$ und mit $(f' \circ T)^+ = (f')^+ \circ T$ und $(f' \circ T)^- = (f')^- \circ T$ folgt die Behauptung.

zu (c): ist eine Übung. $\square$

In der Wahrscheinlichkeitstheorie ergeben sich aus diesem Kapitel sowie aus Kapitel 4 die folgenden Sprechweisen:

5.26 Bemerkungen

(a) Jedes Zufallsexperiment lässt sich mittels einer Zufallsvariablen beschreiben:

$$(\Omega, \mathcal{A}, P), \ X \text{ identische Abbildung}, \ P^X = P \ .$$

Die genaue Angabe von $(\Omega, \mathcal{A}, P)$ tritt in den Hintergrund. Ein Würfelwurf ist zum Beispiel durch irgendeine Zufallsgröße

$$X : (\Omega, \mathcal{A}, P) \to (\{1, \ldots, 6\}, \mathcal{P}(\{1, \ldots, 6\}), P^X)$$

mit $P^X = \frac{1}{6} \sum_{i=1}^{6} \delta_i$ beschrieben (fair!).

(b) Mit der Bezeichung $[X \leq t], [X = Y], \ldots$ für $\{\omega : X(\omega) \leq t\}, \{\omega : X(\omega) = Y(\omega)\}, \ldots$ und $[X \in A]$ für $X^{-1}(A)$ schreiben wir kurz $P(X \leq t), P(X = Y), P(X \in A)$.

(c) Ist X eine Zufallsgröße auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$ und ist $h : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ Borel–messbar, so ist auch $h(X)$ eine Zufallsgröße, etwa $|X|, |X|^p, p \in \mathbb{N}, e^X$ usw. Siehe auch Satz 4.5.

5.27 Definition Es sei X eine Zufallsgröße auf einem W-Raum $(\Omega, \mathcal{A}, P)$. Ist dann $X \geq 0$ oder X P-integrierbar, so heißt

$$\mathbb{E}(X) := \mathbb{E}_P(X) := \int X \, dP \left(= \int_\Omega X \, dP \right)$$

der *Erwartungswert* von X (bzgl. P).

5.28 Definition Ein W-Maß auf $(\mathbb{R}^d, \mathcal{A})$ für $d \geq 1$ und eine beliebige σ-Algebra $\mathcal{A}$ wird als d-dimensionale *Wahrscheinlichkeitsverteilung* (W-Verteilung) bezeichnet. Ein Borel-messbares $f : \mathbb{R}^d \to \mathbb{R}^+$ heißt (d-dimensionale) *Dichtefunktion* oder auch *W-Dichte*, wenn

$$\int_{\mathbb{R}^d} f \, d\lambda^d = 1$$

gilt.

Dann liefert nach Satz 5.25(c)

$$\mathcal{B}^d \ni A \mapsto P(A) := \int_A f \, d\lambda^d$$

ein W-Maß, denn $\int_{\mathbb{R}^d} f\,d\lambda^d = 1$.

Die Verteilung P^X einer reellen Zufallsgröße ist ein W-Maß auf $\mathcal{B}$. Es gilt nach Satz 5.25:

$$\mathbb{E}(f \circ X) = \int f\,dP^X,$$

anders geschrieben $\mathbb{E}_P(f \circ X) = \mathbb{E}_{P^X}(f)$. Hier ist f als BOREL-messbar und nicht-negativ oder als P^X-integrierbar angenommen. Ist also $X \geq 0$ oder X P-integrierbar und wählt man für f die Funktion $x \mapsto x$, so folgt

$$\mathbb{E}(X) = \int x\,dP^X(x).$$

Der Erwartungswert ist nur von der Verteilung von X abhängig! Die Integrierbarkeit von X ist äquivalent zur P^X-Integrierbarkeit von $x \mapsto x$ auf $\mathbb{R}$.

5.29 Beispiele

(a) Da

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2}\,dx = 1 \quad \text{(siehe Analysis III)}\ ,$$

folgt durch Substituition, dass

$$g_{a,\sigma^2}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

für jede Wahl von $a \in \mathbb{R}$ und $\sigma^2 > 0$ eine W-Dichte auf $\mathbb{R}$ bezüglich λ^1 ist:

$$N(a,\sigma^2) := g_{a,\sigma^2} \cdot \lambda^1$$

ist ein W-Maß auf $\mathcal{B}$. Man nennt dies die *Normal-* oder GAUSS-*Verteilung* auf $\mathbb{R}$ zu den Parametern a und σ^2. $N(0,1)$ heißt *standardisierte Normalverteilung* (siehe Abb. 5.1).

(b) Die Funktion

$$x \mapsto \frac{\alpha}{\pi}(\alpha^2 + x^2)^{-1} =: c_\alpha(x)$$

ist für jedes $\alpha > 0$ eine W-Dichte auf $\mathbb{R}$ (bezüglich λ^1), denn

$$\int_{\mathbb{R}} (1 + x^2)^{-1}\,dx = \lim_{n\to\infty} \big[\arctan x\big]_{-n}^{+n} = \pi\ .$$

$\gamma_\alpha := c_\alpha\,\lambda^1$ heißt CAUCHY-*Verteilung zum Parameter* $\alpha > 0$.

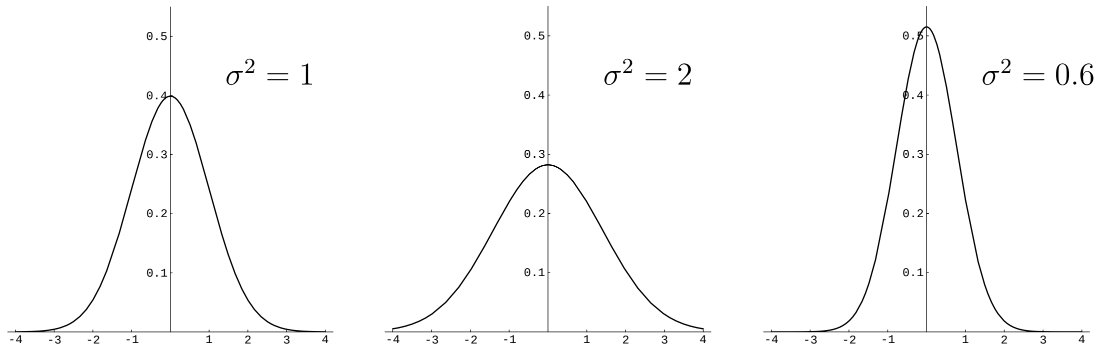


Abb. 5.1: Verschiedene Gaussdichten mit $a = 0$.

5.30 Definition

(a) Das W-Maß auf $(\mathbb{R}^n, \mathcal{B}^n)$, das durch die Dichte

$$\varphi(x_1, \ldots, x_n) := (2\pi)^{-n/2} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\Big), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n ,$$

definiert wird, heißt *Standardnormalverteilung* auf $\mathbb{R}^n$ (siehe Abb. 5.2).

(b) Ein W-Maß P auf $(\mathbb{R}^n, \mathcal{B}^n)$ heißt *Normalverteilung*, wenn eine $n \times n$-Matrix A und $b \in \mathbb{R}^n$ existieren, so dass $P = P_{\mathrm{st}} \phi^{-1}$ ist, wobei $\phi : \mathbb{R}^n \to \mathbb{R}^n$ die affine Abbildung $x \mapsto \phi(x) := Ax + b$ und P_{st} die Standardnormalverteilung sind.

5.31 Satz *Das W-Maß P der obigen Definition besitzt genau dann eine Dichte, wenn A eine invertierbare Matrix ist. In diesem Fall ist die Dichte gegeben durch*

$$\varphi(x, b, \Sigma) := \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\Big(-\frac{1}{2}(x-b)^t \Sigma^{-1}(x-b)\Big), \quad x \in \mathbb{R}^n ,$$

mit $\Sigma = AA^t$. A^t bezeichnet hierbei die Transponierte von A.

Beweis: A sei invertierbar, dann ist ϕ invertierbar. Es gilt für $B \in \mathcal{B}^n$

$$\begin{aligned} P(B) &= P_{\mathrm{st}}(\phi^{-1}(B)) \\ &= \int_{\phi^{-1}(B)} \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2}\, \lambda^n(dx) \\ &= \int_{\mathbb{R}^n} 1_B(\phi(x)) \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}|\phi^{-1}\phi(x)|^2\right) \lambda^n(dx) \\ &= \int_{\mathbb{R}^n} 1_B(y) \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}|\phi^{-1}(y)|^2\right) (\lambda^n \phi^{-1})(dy). \end{aligned}$$

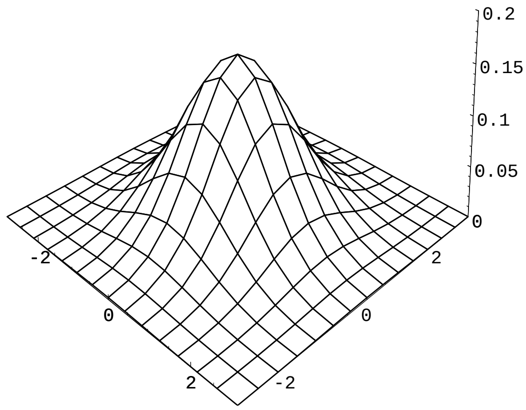


Abb. 5.2: Zweidimensionale Standardnormalverteilung.

Die letzte Gleichheit folgt mittels Satz 5.25 (a). Für bijektive, affine Abbildungen wissen wir mit Satz 3.13

$$\lambda^n(\phi^{-1}(M)) = |\det \phi^{-1}|\lambda^n(M) , \quad M \in \mathcal{B}^n .$$

Also ist $|\det \phi^{-1}| = (\det \Sigma)^{-1/2}$ eine konstante Dichte und mit Satz 5.25(c) folgt

$$\begin{aligned} P(B) &= \int_{\mathbb{R}^n} 1_B(y)(2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}|\phi^{-1}(y)|^2\right) \lambda^n(dy) \\ &= \int_B \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(y-b)^t\Sigma^{-1}(y-b)\right) \lambda^n(dy) , \end{aligned}$$

was zu zeigen war.

Ist nun ϕ nicht invertierbar, so ist $\lambda^n\{\phi(x), x \in \mathbb{R}^n\} = 0$ (siehe Beispiel 2.16(b)), aber $P(\{\phi(x), x \in \mathbb{R}^n\}) = 1$. Also kann P keine Dichte bzgl. λ^n besitzen. $\square$

KAPITEL 6

Fast überall bestehende Eigenschaften

Das μ-Integral wird sich als unempfindlich gegenüber Abänderungen des Integranden auf μ-Nullmengen erweisen, solange der Integrand messbar bleibt. Um diese Eigenschaft formulieren zu können, hat sich der von LEBESGUE eingeführte Begriff *fast überall* als zweckmäßig erwiesen. Im folgenden sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

6.1 Definition Es sei E eine Eigenschaft, die für die Punkte aus Ω entweder richtig oder falsch ist. Man sagt, dass E *μ-fast überall* gilt (dass E für μ-fast alle $x \in \Omega$ richtig ist), wenn es eine μ-Nullmenge N gibt, so dass $E(x)$ für jedes $x \in N^c$ richtig ist. Abkürzend schreiben wir dafür E gilt μ–f.ü.

6.2 Beispiele

(a) Für $f, g\colon \Omega \to \mathbb{R}$ gilt $f \geq g$ μ–f.ü. genau dann, wenn es eine μ-Nullmenge N gibt mit $f(x) \geq g(x)$ für jedes $x \in N^c$.

(b) Es seien $f_j, f \in \mathbb{R}^\Omega$ für $j \in \mathbb{N}$. Dann konvergiert $(f_j)_j$ genau dann μ–f.ü. gegen f, wenn es eine μ-Nullmenge N gibt mit $f_j(x) \to f(x)$ für $x \in N^c$.

(c) $f\colon \Omega \to \mathbb{R}$ ist genau dann μ–f.ü. beschränkt, wenn es eine μ-Nullmenge N und ein $S \geq 0$ gibt mit $|f(x)| \leq S$ für jedes $x \in N^c$.

Es wird *nicht* gefordert, dass die Ausnahmemenge M der $x \in \Omega$, welche die Eigenschaft E nicht haben, zu $\mathcal{A}$ gehört. Es wird nur gefordert, dass M Teilmenge einer geeigneten μ-Nullmenge $N \in \mathcal{A}$ ist.

6.3 Satz

(a) *Für alle $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ gilt:*

$$\int_\Omega f\, d\mu = 0 \Leftrightarrow f = 0 \ \ \mu\text{–f.ü.}$$

(b) *Jedes $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ ist über eine beliebige μ-Nullmenge N integrierbar und es gilt*

$$\int_N f\, d\mu = 0.$$

Beweis: zu (a): Da f messbar ist, liegt $A := [f \neq 0] = [f > 0]$ in $\mathcal{A}$. Wir zeigen

$$\int_\Omega f\, d\mu = 0 \Leftrightarrow \mu(A) = 0 \ .$$

Es sei $\int_\Omega f\, d\mu = 0$. Es seien $A_n := [f > \frac{1}{n}]$, $n \in \mathbb{N}$, also $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, und $A_n \uparrow A$. Aus $\frac{1}{n} 1_{A_n} \leq f$ folgt

$$0 \leq \frac{1}{n}\, \mu(A_n) = \int_\Omega \frac{1}{n}\, 1_{A_n}\, d\mu \leq \int_\Omega f\, d\mu = 0,$$

also $\mu(A_n) = 0$, $n \in \mathbb{N}$, und somit ist $\mu(A) = \lim_{n\to\infty} \mu(A_n) = 0$.

Ist umgekehrt $\mu(A) = 0$, so ist mit $f \leq \infty \cdot 1_A$

$$0 \leq \int_\Omega f\, d\mu \leq \infty \cdot \int_\Omega 1_A\, d\mu = \infty \cdot \mu(A) = 0,$$

also folgt $\int_\Omega f\, d\mu = 0$.

zu (b): Für $f \geq 0$ ist $f\, 1_N \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ und μ–f.ü. Null, also ist nach (a) $\int_N f\, d\mu = \int_\Omega f\, 1_N\, d\mu = 0$. Für ein $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ folgt die Behauptung dann durch Anwendung des Gezeigten auf f^+ und f^-. □

Der erste Teil von Korollar 5.19 liest sich nun so:

6.4 Korollar *Jede μ-integrierbare Funktion $f\colon \Omega \to \overline{\mathbb{R}}$ ist μ–f.ü. auf Ω reellwertig.*

In Kapitel 5 haben wir den $\mathbb{R}$-Vektorraum $\mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ der reellwertigen μ-integrierbaren Funktionen eingeführt. Wir versehen diesen Raum nun mit einer Halbnorm:

6.5 Definition Ist V ein Vektorraum über $\mathbb{K}$, so heißt $\|\cdot\|\colon V \to \mathbb{R}$ eine *Halbnorm* auf V, falls für alle $x, y \in V$ und $\alpha \in \mathbb{K}$ gilt:

(a) $\|x\| \geq 0$

(b) $\|\alpha x\| = |\alpha|\, \|x\|$

(c) $\|x + y\| \leq \|x\| + \|y\|$.

Gilt $\|x\| = 0$ nur für $x = 0$, so heißt $\|\cdot\|$ eine Norm.

6.6 Satz *Die Menge $\mathcal{L}^1$ ist ein $\mathbb{R}$-Vektorraum mit Halbnorm*

$$\|f\|_1 := \int_\Omega |f|\, d\mu, \ \ f \in \mathcal{L}^1.$$

Die Abbildung $I : \mathcal{L}^1 \to \mathbb{R}$, $I(f) := \int_\Omega f\,d\mu$, ist eine positive Linearform auf $\mathcal{L}^1$ mit der Eigenschaft

$$|I(f) - I(g)| \leq \|f - g\|_1, \quad f, g \in \mathcal{L}^1.$$

Beweis: In Satz 5.20 und Bemerkung 5.22 ist gezeigt worden, dass $\mathcal{L}^1$ ein $\mathbb{R}$-Vektorraum ist und dass I eine Linearform ist. Für $f \in \mathcal{L}^1$ mit $f \geq 0$ gilt $I(f) \geq 0$. $\|\cdot\|_1$ ist eine Halbnorm, denn $\int |f|\,d\mu \geq 0$, und die anderen Eigenschaften sind unmittelbar klar. Weiter ist

$$\left| \int_\Omega (f-g)\,d\mu \right| \leq \int_\Omega |f-g|\,d\mu = \|f-g\|_1$$

für $f, g \in \mathcal{L}^1$. □

6.7 Bemerkung $\|\cdot\|_1$ ist nicht notwendig eine Norm. Denn mit Satz 6.3 (a) gilt $\|f\|_1 = 0$ genau dann, wenn $[|f| > 0]$ eine μ-Nullmenge ist. Gibt es also eine nicht-leere Nullmenge $A \in \mathcal{A}$, so ist $f := 1_A \in \mathcal{L}^1$, $f \neq 0$, aber $\|f\|_1 = 0$.

6.8 Satz *Es seien $f, g \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$, die μ–f.ü. auf Ω gleich sind. Dann gilt:*

(a) *Ist $f \geq 0$ und $g \geq 0$, so folgt $\int f\,d\mu = \int g\,d\mu$.*

(b) *Ist f μ-integrierbar, so ist g μ-integrierbar und $\int f\,d\mu = \int g\,d\mu$.*

Beweis: zu (a): Es sei $N := [f \neq g]$. Also ist $N \in \mathcal{A}$ und $\mu(N) = 0$ nach Voraussetzung. Somit ist $\int_N f\,d\mu = \int_N g\,d\mu = 0$ nach Satz 6.3 (b). Mit $M := N^c$ ist dann

$$\int_M f\,d\mu = \int_\Omega 1_M\,f\,d\mu = \int_\Omega 1_M\,g\,d\mu = \int_M g\,d\mu.$$

Also folgt die Behauptung, da $\int_{M \cup N} f\,d\mu = \int_M f\,d\mu + \int_N f\,d\mu$ für $M \cap N = \emptyset$ (siehe Satz 5.3 und die einfache Ausweitung auf μ-integrierbare Funktionen).

zu (b): Es ist $f^+ = g^+$ μ–f.ü. und $f^- = g^-$ μ–f.ü. Also folgt mit (a) $\int f^+\,d\mu = \int g^+\,d\mu$ und $\int f^-\,d\mu = \int g^-\,d\mu$. Da f μ-integrierbar ist, sind beide Integrale nichtnegative reelle Zahlen. Daraus folgt die Behauptung. □

6.9 Korollar *Es seien $f, g \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ und $|f| \leq g$ μ–f.ü. Dann ist mit g auch f μ-integrierbar.*

Beweis: Wir setzen $g' := g \vee |f|$, dann ist g' messbar und $g' = g$ μ–f.ü. und $|f| \leq g'$. Nach Satz 6.8 ist g' somit μ-integrierbar und nach Satz 5.17 dann auch f. □

Es sei N eine μ-Nullmenge und f eine auf N^c definierte, $N^c \cap \mathcal{A}$-messbare Funktion. f heißt μ*–f.ü. definierte*, $\mathcal{A}$-messbare Funktion. f kann zu einer $\mathcal{A}$-messbaren Funktion auf Ω fortgesetzt werden: Man setze

$$f_{N^c}(x) := \begin{cases} f(x)\,, & x \in N^c \\ 0\,, & x \in N. \end{cases}$$

Jede andere Fortsetzung von f auf Ω ist dann μ–f.ü. gleich f_{N^c}. Nach Satz 6.8 sind somit alle Fortsetzungen μ-integrierbar oder keine der Fortsetzungen ist μ-integrierbar. Sind sie μ-integrierbar, so haben alle dasselbe μ-Integral. Dies führt zu der folgenden Definition:

6.10 Definition Es sei f eine μ–f.ü. auf Ω definierte, $\mathcal{A}$-messbare, numerische Funktion. Sie heißt μ*-integrierbar*, wenn sie zu einer μ-integrierbaren, auf ganz Ω definierten Funktion f' fortgesetzt werden kann. $\int f'\,d\mu$ heißt dann das μ-Integral von f. Es wird auch mit $\int f\,d\mu$ bezeichnet.

6.11 Bemerkung Es seien f, g zwei μ-integrierbare Funktionen. Nach Korollar 6.4 ist $\mathcal{A} \ni N := \{|f| = \infty\} \cup \{|g| = \infty\}$ eine μ-Nullmenge, also ist $f + g$ auf N^c definiert, also im Sinne der Definition 6.10 integrierbar. Es gilt $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$ (siehe Satz 5.20). Die zusätzliche Bemerkung, $f+g$ sei überall wohldefiniert, die eigentlich bei der Betrachtung einer Summe zweier messbarer numerischer Funktionen erfolgen muß, kann also weggelassen werden.

Wir betrachten nun einen zentralen *Vertauschungssatz* für Integrale und Grenzwerte, den Satz von Lebesgue über die majorisierte Konvergenz:

6.12 Satz (von der majorisierten Konvergenz, Lebesgue) *Die Funktionen $(f_j)_j$ und f seien in $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ und $\lim_{j\to\infty} f_j = f$ μ–f.ü. Ferner gebe es ein $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ mit $|f_j| \le g$ μ–f.ü., $j \in \mathbb{N}$. Dann sind f und f_j, $j \in \mathbb{N}$, μ-integrierbar und es gelten*

$$\lim_{j\to\infty} \int f_j\,d\mu = \int f\,d\mu$$

und

$$\lim_{j\to\infty} \int |f_j - f|\,d\mu = 0.$$

Beweis: Nach Korollar 6.9 sind f und alle f_j, $j \in \mathbb{N}$, μ-integrierbar. Nach den μ–fast überall Resultaten (siehe Korollar 6.4, Satz 6.8) haben ohne Einschränkung der Allgemeinheit f und g und alle f_j, $j \in \mathbb{N}$, Werte in $\mathbb{R}$ (überall) und überall gilt $\lim_{j\to\infty} f_j = f$, $|f_j| \le g$, $j \in \mathbb{N}$. Dann ist

$$g_j := |f| + g - |f_j - f|$$

in $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}}^+)$ und das Lemma von FATOU (Satz 5.14) liefert

$$\begin{aligned}\int \big(|f| + g\big)\, d\mu = \int \big(\liminf_j g_j\big)\, d\mu \;&\le\; \liminf_j \int g_j\, d\mu \\ &=\; \int \big(|f| + g\big)\, d\mu - \limsup_j \int |f_j - f|\, d\mu.\end{aligned}$$

Hier ist das Integral $\int \big(|f| + g\big)\, d\mu$ endlich, also folgt $\lim_{j\to\infty} \int |f_j - f|\, d\mu = 0$ und somit liefert $|\int f_j\, d\mu - \int f\, d\mu| \le \int |f_j - f|\, d\mu$ die Behauptung. □

Wir haben somit bereits *die drei wichtigsten* Konvergenzsätze der Integrationstheorie kennengelernt: Satz 5.10 *über die monotone Konvergenz*, Satz 5.14, das *Lemma von* FATOU, und Satz 6.12 *von der majorisierten Konvergenz*. Die LEBESGUEsche Integrationstheorie ermöglicht sehr allgemeine und flexible Kriterien für die Vertauschbarkeit von Grenzwerten und Integralen. So ist in Satz 6.12 nur punktweise Konvergenz vorausgesetzt. Bei den CAUCHY-RIEMANN-Integralen wird f_j auf $[a, b] \subset \mathbb{R}$ definiert und für $f_j \to f$ die *gleichmäßige* Konvergenz gegen f gefordert, um einen Vertauschungs-Konvergenzsatz erhalten zu können.

Gegeben sei der Maßraum $(\mathbb{R}^d, \mathcal{B}^d, \lambda^d)$. Entlang der Definition 5.15 spricht man für Funktionen aus $\mathcal{L}_0(\mathbb{R}^d, \mathcal{B}^d, \overline{\mathbb{R}})$ von LEBESGUE *integrierbaren* Funktionen und von deren LEBESGUE *Integral*. Tatsächlich ist jede Regelfunktion auf einem abgeschlossenen Intervall in $\mathbb{R}$ (sprungstetige Funktion) BOREL-messbar und das CAUCHY-RIEMANN Integral stimmt mit dem LEBESGUE Integral überein. Damit stehen im Rahmen des LEBESGUE Integrals die Methoden zur Verfügung, die für das CAUCHY-RIEMANN Integral entwickelt wurden. Weiter wird gezeigt, dass eine beschränkte, reelle Funktion auf einem kompakten Intervall genau dann RIEMANN integrierbar ist, wenn die Menge der Unstetigkeitsstellen eine λ^1-Nullmenge ist. Es folgt daraus, dass das LEBESGUEsche Integral eine echte Erweiterung des RIEMANNschen, und daher des CAUCHY-RIEMANNschen Integrals ist.

Es sei fortan $X \subset \mathbb{R}^d$ mit $\lambda^d(X) > 0$ gegeben. Zur Abkürzung setzen wir

$$\mathcal{L}^1(X) := \mathcal{L}^1(X, \mathcal{B}^d|_X, \lambda^d|_X).$$

6.13 Satz *Es seien $-\infty \le a < b \le \infty$ und $f\colon (a, b) \to \mathbb{R}$ sei absolut integrierbar im Sinne des* CAUCHY-RIEMANN *Integrals, d.h. es existiere $\int_a^b |f|$ in $\mathbb{R}$ (Notation für das* CAUCHY-RIEMANN *Integral). Dann ist $f \in \mathcal{L}^1((a, b))$ und*

$$\int_{(a,b)} f\, d\lambda^1 = \int_a^b f.$$

Wir geben den Beweis an, werden ihn aber in der Vorlesung nicht besprechen.

Beweis: 1. Schritt: Es seien zunächst $a < \alpha < \beta < b$. Ist $g\colon [\alpha, \beta] \to \mathbb{R}$ eine Treppenfunktion, so ist g $\mathcal{B}$-einfach und

$$\int_{(\alpha,\beta)} g \, d\lambda^1 = \int_\alpha^\beta g. \tag{6.1}$$

Es sei nun $g\colon [\alpha, \beta] \to \mathbb{R}$ eine Regelfunktion. Dann existiert eine Folge $(g_j)_j$ von Treppenfunktionen, die gleichmäßig gegen g konvergiert. Also ist g Borel-messbar und $g \in \mathcal{L}^1((\alpha, \beta))$, denn g ist beschränkt (bekannt aus der Analysis), also ist $|g| \le \gamma$ für ein $\gamma \in [0, \infty)$, also $\int_{(\alpha,\beta)} |g| \, d\lambda^1 \le \gamma(\beta - \alpha)$. Da die Folge $(g_j)_j$ gleichmäßig konvergiert und g beschränkt ist, gibt es ein $M \ge 0$ mit $|g_j| \le M$ für alle $j \in \mathbb{N}$. Nach Satz 6.12 folgt also

$$\lim_{j\to\infty} \int_{(\alpha,\beta)} g_j \, d\lambda^1 = \int_{(\alpha,\beta)} g \, d\lambda^1.$$

Nach der Definition des Cauchy-Riemann Integrals und (6.1) folgt

$$\int_\alpha^\beta g = \lim_{j\to\infty} \int_\alpha^\beta g_j = \lim_{j\to\infty} \int_{(\alpha,\beta)} g_j \, d\lambda^1 = \int_{(\alpha,\beta)} g \, d\lambda^1.$$

2. Schritt: Es sei nun $c \in (a, b)$ und $(\beta_j)_j$ eine Folge in (c, b) mit $\beta_j \uparrow b$. Sei $g := 1_{[c,b)} f$ und $g_j := 1_{[c,\beta_j]} f$, $j \in \mathbb{N}$. Nach dem 1. Schritt ist g_j in $\mathcal{L}^1(\mathbb{R})$ für jedes $j \in \mathbb{N}$. Die Folge $(g_j)_j$ konvergiert punktweise gegen g und die Folge $(|g_j|)_j$ konvergiert wachsend gegen $|g|$. Somit ist g Borel-messbar und mit Hilfe des 1. Schritts folgt

$$\int_\mathbb{R} |g_j| \, d\lambda^1 = \int_{(c,\beta_j)} |f| \, d\lambda^1 = \int_c^{\beta_j} |f|.$$

Die absolute Konvergenz von $\int_c^b f$ liefert

$$\lim_{j\to\infty} \int_\mathbb{R} |g_j| \, d\lambda^1 = \lim_{j\to\infty} \int_c^{\beta_j} |f| = \int_c^b |f|.$$

Der Satz von der monotonen Konvergenz (Satz 5.10) liefert

$$\int_\mathbb{R} |g| \, d\lambda^1 = \lim_{j\to\infty} \int_\mathbb{R} |g_j| \, d\lambda^1,$$

also folgt $g \in \mathcal{L}^1(\mathbb{R})$. Der Satz von der majorisierten Konvergenz (Satz 6.12) kann somit auf die Folge $(g_j)_j$ angewendet werden:

$$\lim_{j\to\infty} \int_\mathbb{R} g_j \, d\lambda^1 = \int_\mathbb{R} g \, d\lambda^1 = \int_{[c,b)} f \, d\lambda^1.$$

Weiter ist nach Schritt 1 $\int_{\mathbb{R}} g_j \, d\lambda^1 = \int_c^{\beta_j} f$, also

$$\lim_{j\to\infty} \int_{\mathbb{R}} g_j \, d\lambda^1 = \lim_{j\to\infty} \int_c^{\beta_j} f = \int_c^b f.$$

Somit stimmen die Grenzwerte $\int_{[c,b)} f \, d\lambda^1$ und $\int_c^b f$ überein. Analog zeigt man, dass $1_{(a,c]} f \in \mathcal{L}^1(\mathbb{R})$ und $\int_{(a,c]} f \, d\lambda^1 = \int_a^c f$ gilt. Somit ist f λ^1-integrierbar mit $\int_{(a,b)} f \, d\lambda^1 = \int_a^b f$. Damit ist der Satz bewiesen. □

6.14 Bemerkung Ist $f\colon (a,b) \to \mathbb{R}$ zulässig (auf jedem kompakten Teilintervall sprungstetig) und $\int_a^b f$ existiere als uneigentliches Integral, so muß f nicht zu $\mathcal{L}^1((a,b))$ gehören. Dazu betrachten wir

$$f(x) := \begin{cases} 0\,, & x \in (-\infty, 0), \\ \frac{(-1)^j}{j}\,, & x \in [j-1, j),\ j \in \mathbb{N}. \end{cases}$$

Dann ist f zulässig und

$$\int_{-\infty}^{\infty} f = \sum_{j=1}^{\infty} (-1)^j/j,$$

also existiert $\int_{-\infty}^{\infty} f$ in $\mathbb{R}$. Angenommen, es ist $f \in \mathcal{L}_1(\mathbb{R})$. Dann gilt $\int_{\mathbb{R}} |f| \, d\lambda^1 < \infty$, aber aus dem Satz über die monotone Konvergenz (Satz 5.10) folgt

$$\int_{\mathbb{R}} |f| \, d\lambda^1 = \lim_{k\to\infty} \int_{\mathbb{R}} 1_{[0,k]} |f| \, d\lambda^1 = \lim_{k\to\infty} \sum_{j=1}^{k} \frac{1}{j} = \infty,$$

und das ist ein Widerspruch.

6.15 Satz *Es sei I ein kompaktes Intervall und $f\colon I \to \mathbb{R}$ sei beschränkt. Genau dann ist f Riemann integrierbar, wenn f λ^1–f.ü. stetig ist. In diesem Fall ist f Lebesgue integrierbar und das Riemannsche Integral stimmt mit dem Lebegueschen überein.*

Wir geben auch hier für das Selbststudium den Beweis an.

Beweis: 1. Schritt: Ohne Einschränkung der Allgemeinheit betrachten wir $I = [0,1]$. Für $k \in \mathbb{N}$ sei $\mathcal{Z}_k := (\zeta_{0,k,\ldots,}\zeta_{2^k,k})$ die Zerlegung von $[0,1]$ mit $\zeta_{j,k} := j2^{-k}$ für $j = 0, \ldots, 2^k$. Ferner seien

$$I_{0,k} := [\zeta_{0,k}, \zeta_{1,k}]\ ,\ I_{j,k} := (\zeta_{j,k}, \zeta_{j+1,k}]\ ,\ j = 1, \ldots, 2^k - 1,$$

$$\alpha_{j,k} := \inf_{x\in\overline{I}_{j,k}} f(x)\ ,\ \beta_{j,k} := \sup_{x\in\overline{I}_{j,k}} f(x),$$

sowie

$$g_k := \sum_{j=0}^{2^k-1} \alpha_{j,k}\, 1_{I_{j,k}}\ ,\ h_k := \sum_{j=0}^{2^k-1} \beta_{j,k}\, 1_{I_{j,k}},\ k \in \mathbb{N}.$$

Dann ist $(g_k)_k$ eine wachsende Folge und $(h_k)_k$ ist eine fallende Folge von $\mathcal{B}$-einfachen Funktionen. Somit sind $g = \lim_k g_k$ und $h = \lim_k h_k$ punktweise definiert und $\mathcal{B}$-messbar und es gilt $g \leq f \leq h$. Es gilt

$$\int_{[0,1]} g_k\, d\lambda^1 = \underline{S}(f,[0,1],\mathcal{Z}_k) \text{ und } \int_{[0,1]} h_k\, d\lambda^1 = \overline{S}(f,[0,1],\mathcal{Z}_k).$$

Hierbei bezeichnet $\underline{S}(f,[0,1],\mathcal{Z}_k)$ bzw. $\overline{S}(f,[0,1],\mathcal{Z}_k)$ die Unter- bzw. Obersumme von f über $[0,1]$ bezüglich der Zerlegung $\mathcal{Z}_k$ (siehe Analysis II). In der Notation des oberen und unteren Riemann Integrals $\overline{\int_I} f$, $\underline{\int_I} f$ (siehe Analysis II) folgt mittels des Satzes über die monotone Konvergenz (Satz 5.10):

$$\int_{[0,1]} (h-g)\, d\lambda_1 = \overline{\int_{[0,1]}} f - \underline{\int_{[0,1]}} f. \tag{6.2}$$

2. Schritt: Es seien $R = \bigcup_{k\in\mathbb{N}} \{\zeta_{0,k,\ldots}\zeta_{2^k,k}\}$ und C die Menge der Stetigkeitspunkte von f. Dann gilt

$$[g = h] \cap R^c \subset C \subset [g = h]. \tag{6.3}$$

Um dies zu beweisen, seien $\epsilon > 0$ und $x_0 \in R^c$ mit $g(x_0) = h(x_0)$. Dann existiert ein $k \in \mathbb{N}$ mit $h_k(x_0) - g_k(x_0) < \epsilon$ und ein $j \in \{0,\ldots,2^{k-1}\}$, so dass x_0 in $(\zeta_{j,k}, \zeta_{j+1,k})$ liegt. Für $x \in I_{j,k}$ ist

$$\mid f(x) - f(x_0) \mid \leq \sup_{y\in \overline{I}_{j,k}} f(y) - \inf_{y\in \overline{I}_{j,k}} f(y) = h_k(x_0) - g_k(x_0) < \epsilon,$$

also ist f in x_0 stetig. Damit ist die Inklusion $[g = h] \cap R^c \subset C$ bewiesen.

Sei weiter $x_0 \in C$ und $\epsilon > 0$. Dann existiert ein $\delta > 0$ mit $\mid f(x) - f(x_0) \mid < \epsilon/2$ für $x \in [x_0 - \delta, x_0 + \delta] \cap [0,1]$. Es sei $k_0 \in \mathbb{N}$ mit $2^{-k_0} \leq \delta$. Man findet für jedes $k \geq k_0$ ein $j \in \{0,\ldots,2^k - 1\}$ mit $x_0 \in I_{j,k} \subset [x_0 - \delta, x_0 + \delta]$. Also

$$0 \leq h_k(x_0) - g_k(x_0) = \sup_{x\in \overline{I}_{j,k}} \big(f(x) - f(x_0)\big) - \inf_{x\in \overline{I}_{j,k}} \big(f(x) - f(x_0)\big) < \epsilon.$$

Somit folgt $h(x_0) - g(x_0) = \lim_k \big(h_k(x_0) - g_k(x_0)\big) = 0$. Damit ist die Inklusion $C \subset [g = h]$ bewiesen und somit ist (6.3) bewiesen.

3. Schritt: Ist f Riemann integrierbar, so ist

$$\underline{\int_{[0,1]}} f = \overline{\int_{[0,1]}} f = \int_{[0,1]} f,$$

also ist $h = g = f$ λ^1–f.ü. nach (6.2). Damit ist f $\mathcal{B}$-messbar. Da f beschränkt ist, folgt $f \in \mathcal{L}^1([0,1])$. Weiter ist $|g_k| \leq \|f\|_\infty$ λ^1–f.ü. für $k \in \mathbb{N}$, und somit ergibt der Satz von der majorisierten Konvergenz (Satz 6.12)

$$\int_{[0,1]} g \, d\lambda^1 = \lim_{k\to\infty} \int_{[0,1]} g_k \, d\lambda^1 = \lim_{k\to\infty} \underline{S}(f, [0,1], \mathcal{Z}_k) = \int_0^1 f.$$

Da $g = f$ λ^1–f.ü., folgt $\int_{[0,1]} f \, d\lambda^1 = \int_0^1 f$. Es bleibt zu zeigen, dass die Menge C^c der Unstetigkeitsstellen von f eine λ^1-Nullmenge ist. Nach (6.3) gilt $C^c \subset [h \neq g] \cup R$. Da aber $g = h$ λ^1–f.ü., und die Menge R abzählbar ist, folgt

$$\lambda^1(C^c) \leq \lambda^1\big([h \neq g] \cup R\big) = 0.$$

Damit bilden die Unstetigkeitsstellen von f eine λ^1-Nullmenge.

4. Schritt: Es sei nun C^c eine λ^1-Nullmenge, dann ist nach (6.3) auch $[g \neq h]$ eine λ^1-Nullmenge und die Riemann Integrierbarkeit von f folgt aus (6.2). Somit ist der Satz bewiesen. □

6.16 Korollar *Das* Lebesgue *Integral ist eine echte Erweiterung des* Riemannschen *Integrals.*

Beweis: Man betrachte die Dirichlet*funktion*

$$f\colon [0,1] \to \mathbb{R} \quad f(x) = \begin{cases} 1\,, & x \in \mathbb{Q}, \\ 0\,, & x \notin \mathbb{Q}. \end{cases}$$

Da $f = 0$ λ^1–f.ü. (siehe Beispiel 2.16 (b)), ist $f \in \mathcal{L}^1([0,1])$, siehe 6.8 (b). Weiter ist f in keinem Punkt stetig, also ist f nach Satz 6.15 nicht Riemann integrierbar. □

KAPITEL 7

Die Lebesgueschen Räume $\mathcal{L}^p$ und L^p und Momente

Gegeben seien der Messraum $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ und eine numerische Funktion $f : \mathbb{N} \to \overline{\mathbb{R}}$. Es sei weiter ϱ ein Maß auf $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ mit $\varrho(\{n\}) = \alpha_n$ (der Fall $\alpha_n = 1$ liefert das Zählmaß). Dann gilt analog zu Beispiel 5.13(b)

$$\int f\, d\varrho = \sum_{n=1}^{\infty} f(n)\, \varrho(\{n\}) = \sum_{n=1}^{\infty} f(n)\, \alpha_n.$$

Ist nun $\alpha_n = n^{-p-1}$, $p \in (1, \infty)$, und $f(n) = n$, so ist f ϱ-integrierbar, aber die p-te Potenz $f^p(n) = n^p$ ist *nicht* ϱ-integrierbar. Insbesondere ist im Fall $p = 2$ die Funktion $f^2 = f \cdot f$ nicht ϱ-integrierbar. Somit ist das Produkt zweier integrierbarer Funktionen im Allgemeinen nicht wieder integrierbar. Wir wollen in diesem Kapitel diejenigen messbaren numerischen Funktionen untersuchen, für welche $|f|^p$ integrierbar ist.

Im folgenden sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

Ist $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$, so ist $|f|^p$, $p \in (0, \infty)$, auch in $\mathcal{L}_0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$, denn für $\alpha \in \mathbb{R}$ ist

$$[|f|^p \geq \alpha] = \begin{cases} [|f| \geq \alpha^{1/p}]\ , & \alpha > 0, \\ \Omega\ , & \text{sonst.} \end{cases}$$

Also liefert Satz 4.3 die Behauptung.

7.1 Definition Es sei $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R})$. Man nennt f *μ-wesentlich beschränkt*, wenn es ein $\alpha \geq 0$ gibt mit $\mu([|f| > \alpha]) = 0$. Dann heißt

$$\|f\|_\infty := \text{ess-sup}_{x\in\Omega}|f(x)| := \inf\{\alpha \geq 0\colon \mu([|f| > \alpha]) = 0\}$$

μ-wesentliches Supremum von f. Wir setzen

$$\|f\|_p := \Big(\int_\Omega |f|^p\, d\mu \Big)^{1/p},\ \ p \in (0, \infty),$$

mit der Vereinbarung $\infty^{1/p} := \infty$. Dann heißt

$$\mathcal{L}^p(\Omega, \mathcal{A}, \mu) := \{f \in \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R}) : \|f\|_p < \infty\},\ \ p \in (0, \infty],$$

Lebesguescher Raum über Ω bezüglich μ. Für $p \in [1, \infty]$ ist der zu p *duale Exponent* gegeben durch

$$p' := \begin{cases} \infty\ , & p = 1 \\ p/(p-1)\ , & p \in (1, \infty) \\ 1\ , & p = \infty. \end{cases}$$

Mit der Festsetzung des dualen Exponenten gilt

$$\frac{1}{p} + \frac{1}{p'} = 1, \;\; p \in [1, \infty].$$

Wir werden zeigen, dass für $p \geq 1$ die Abbildung $\| \cdot \|_p : \mathcal{L}^p(\Omega, \mathcal{A}, \mu) \to \mathbb{R}$ eine Halbnorm auf dem $\mathbb{R}$-Vektorraum $\mathcal{L}^p := \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ ist und das $(\mathcal{L}^p, \| \cdot \|_p)$ vollständig ist. Dabei werden wir noch klären, was Vollständigkeit für einen Vektorraum, versehen mit einer Halbnorm, bedeutet.

7.2 Bemerkungen

(a) Eine Funktion $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R})$ ist genau dann μ-wesentlich beschränkt, wenn $\|f\|_\infty < \infty$.

(b) Für $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R})$ gilt $|f| \leq \|f\|_\infty$ μ–f.ü.. Denn im Fall $\|f\|_\infty = \infty$ ist dies klar. Gilt $\|f\|_\infty < \infty$, so ist $[|f| > \|f\|_\infty + 2^{-j}]$ für jedes $j \in \mathbb{N}$ eine μ-Nullmenge, also auch $[|f| > \|f\|_\infty] = \bigcup_{j \in \mathbb{N}} [|f| > \|f\|_\infty + 2^{-j}]$.

(c) Es seien f und g μ-wesentlich beschränkt und $\alpha \in \mathbb{R}$. Dann ist auch $\alpha\, f + g$ μ-wesentlich beschränkt und es gilt

$$\|\alpha\, f + g\|_\infty \leq |\alpha|\, \|f\|_\infty + \|g\|_\infty.$$

Denn gemäß (b) gibt es μ-Nullmengen M und N mit $|f(x)| \leq \|f\|_\infty$ für $x \in M^c$ und $|g(x)| \leq \|g\|_\infty$ für $x \in N^c$. Also gilt

$$|\alpha f(x) + g(x)| \leq |\alpha|\, \|f\|_\infty + \|g\|_\infty, \;\; x \in (M \cup N)^c = M^c \cap N^c.$$

7.3 Satz *$\mathcal{L}^p$ ist ein reeller Vektorraum für $1 \leq p \leq \infty$ (ein Untervektorraum von $\mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R})$).*

Beweis: Für $p = \infty$ folgt dies aus Bemerkung 7.2 (c). Der Fall $p = 1$ wurde in Satz 5.20 bereits betrachtet. Es sei $p \in (1, \infty)$. Betrachte

$$|a + b|^p \leq \big(2(|a| \vee |b|)\big)^p \leq 2^p \left(|a|^p + |b|^p\right), \;\; a, b \in \mathbb{R}.$$

Dann ist mit $f, g \in \mathcal{L}^p$

$$\int_\Omega |f + g|^p \, d\mu \leq 2^p \left(\int_\Omega |f|^p \, d\mu + \int_\Omega |g|^p \, d\mu \right) < \infty,$$

also $\|f + g\|_p < \infty$. $\alpha f \in \mathcal{L}^p$ für $\alpha \in \mathbb{R}$ ist klar. □

7.4 Satz (Höldersche Ungleichung) *Es sei $p \in [1, \infty]$. Für $f \in \mathcal{L}^p$ und $g \in \mathcal{L}^{p'}$ ist $f\, g \in \mathcal{L}^1$ und*

$$\left| \int_\Omega f\, g \, d\mu \right| \leq \int_\Omega |f\, g| \, d\mu \leq \|f\|_p \, \|g\|_{p'}.$$

Im Fall $p = p' = 2$ heißt die Ungleichung die CAUCHY-SCHWARZ*sche Ungleichung.*

7.5 Korollar *Wähle für μ das Zählmaß ϱ auf* $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, *so folgt*

$$\sum_{n\geq 1} |x_n y_n| \leq \left(\sum_{n\geq 1} |x_n|^p\right)^{1/p} \left(\sum_{n\geq 1} |y_n|^{p'}\right)^{1/p'}$$

für $x_n, y_n \in \mathbb{R}$, $n \in \mathbb{N}$.

Beweis von Satz 7.4: Wir betrachten zuerst den Fall $p = 1$. Nach Bemerkung 7.2 (b) existiert eine μ-Nullmenge N mit $|g(x)| \leq \|g\|_\infty$, $x \in N^c$. Dann folgt

$$\int_{N^c} |f\, g|\, d\mu \leq \|g\|_\infty \int_{N^c} |f|\, d\mu = \|g\|_\infty \, \|f\|_1 < \infty$$

nach Satz 6.8 und Satz 5.20. Dann ist $f\, g$ in $\mathcal{L}^1$ nach Satz 6.8 und Satz 5.17. Weiter folgt mit Satz 5.20

$$\Big|\int_\Omega f\, g\, d\mu\Big| \leq \int_\Omega |f\, g|\, d\mu = \int_{N^c} |f\, g|\, d\mu \leq \|f\|_1 \, \|g\|_\infty.$$

Der Fall $p = \infty$ wird analog zu $p = 1$ behandelt.

Es sei nun $p \in (1, \infty)$: Für $f = 0$ μ–f.ü. oder $g = 0$ μ–f.ü. verschwindet $f\, g$ μ–f.ü. und Satz 6.3 (a) liefert die Behauptung. Sonst gilt $\|f\|_p > 0$ und $\|g\|_{p'} > 0$ (man verwende $\|f\|_1 = 0 \Leftrightarrow f = 0$ μ–f.ü. für $f \in \mathcal{L}^1$, Satz 6.3 (a)). Nun gilt die folgende Ungleichung (Übung oder aus der Analysis bekannt):

$$\xi\, \eta \leq \frac{1}{p}\, \xi^p + \frac{1}{p'}\, \eta^{p'}$$

für alle $\xi, \eta \in [0, \infty]$. Setzen wir nun $\xi := \frac{|f|}{\|f\|_p}$ und $\eta := \frac{|g|}{\|g\|_{p'}}$, so folgt

$$\frac{|f\, g|}{\|f\|_p \, \|g\|_{p'}} \leq \frac{1}{p}\, \frac{|f|^p}{\|f\|_p^p} + \frac{1}{p'}\, \frac{|g|^{p'}}{\|g\|_{p'}^{p'}}\ ,$$

also

$$\int_\Omega |f\, g|\, d\mu \leq \frac{1}{p}\, \|f\|_p^{1-p}\, \|g\|_{p'} \int_\Omega |f|^p\, d\mu + \frac{1}{p'}\, \|f\|_p \, \|g\|_{p'}^{1-p'} \int_\Omega |g|^{p'}\, d\mu = \|f\|_p \, \|g\|_{p'},$$

und somit $f\, g \in \mathcal{L}^1$ mit Satz 5.17 und

$$\Big|\int_\Omega f\, g\, d\mu\Big| \leq \|f\, g\|_1 \leq \|f\|_p \, \|g\|_{p'}.$$

$\square$

Die nun folgende Ungleichung von MINKOWSKI bringt zum Ausdruck, dass $\|\cdot\|_p$ für $1 \le p \le \infty$ der Dreiecks-Ungleichung genügt.

7.6 Satz (Minkowskische Ungleichung) *Es sei $p \in [1,\infty]$. Für $f, g \in \mathcal{L}^p$ ist $f + g \in \mathcal{L}^p$ und*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p.$$

Beweis: Der Fall $p = 1$ folgt mit Satz 6.6. Der Fall $p = \infty$ folgt mit Bemerkung 7.2 (c). Es sei fortan $p \in (1,\infty)$. Auf Grund der Äquivalenz

$$|f+g|^{p-1} \in \mathcal{L}^{p'} \Leftrightarrow |f+g| \in \mathcal{L}^p$$

folgt aus der HÖLDERschen Ungleichung

$$\int_\Omega |h| \cdot |f+g|^{p-1}\, d\mu \le \|h\|_p \cdot \|\, |f+g|^{p-1}\, \|_{p'} = \|h\|_p \cdot \|f+g\|_p^{p/p'}$$

für $h \in \mathcal{L}^p$, also

$$\int_\Omega |f+g|^p\, d\mu \le \int_\Omega |f| \cdot |f+g|^{p-1}\, d\mu + \int_\Omega |g| \cdot |f+g|^{p-1}\, d\mu \le \big(\|f\|_p + \|g\|_p\big) \cdot \|f+g\|_p^{p/p'} .$$

Es ist $\|f+g\|_p < \infty$ und $p/p' = p - 1$, und somit folgt die Behauptung. □

7.7 Satz *Für $p \in [1,\infty]$ ist $\mathcal{L}^p$ ein $\mathbb{R}$-Vektorraum und $\|\cdot\|_p$ eine Halbnorm auf $\mathcal{L}^p$.*

Beweis: Die Eigenschaft, $\mathbb{R}$-Vektorraum zu sein, ist in Satz 7.3 bewiesen worden. Im Fall $p = 1$ folgt aus Satz 6.6, dass $\|\cdot\|_1$ eine Halbnorm ist. Der Fall $p \in (1,\infty]$ folgt aus der MINKOWSKIschen Ungleichung. □

7.8 Bemerkung Es sei $\mathcal{N} = \{f \in \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R}) \colon f = 0\ \mu$–f.ü.$\}$. Dann sind für $f \in \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R})$ äquivalent:

(a) $\|f\|_p = 0$ für alle $p \in [1,\infty]$,

(b) $\|f\|_p = 0$ für ein $p \in [1,\infty]$,

(c) $f \in \mathcal{N}$.

Beweis: (a) $\Rightarrow$ (b): ist klar.

(b) $\Rightarrow$ (c): zu zeigen ist $\|f\|_1 = 0$ (denn dies ist äquivalent zu $f = 0$ μ–f.ü., siehe Satz 6.3 (a)). Aber mit $|f| \le \|f\|_\infty$ μ-fast überall (siehe Bemerkung 7.2(b)) ist dies klar.

(c) $\Rightarrow$ (a): Für $p \in [1,\infty)$ ist dies mit Satz 6.8 (b) klar. Der Fall $p = \infty$ ist klar. □

7.9 Bemerkung Es sei μ ein endliches Maß. Für $p \in [1, \infty]$ ist dann

$$\mathcal{EF}(\Omega, \mathbb{R}) \subset \mathcal{L}^p(\Omega, \mathcal{A}, \mu) \subset \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R})$$

im Sinne von Untervektorräumen.

Beweis: Jede $\mathcal{A}$-einfache Funktionen ist μ-wesentlich beschränkt. Es sei $p \in [1, \infty)$, und $\varphi = \sum_{j=1}^m \alpha_j 1_{A_j} \in \mathcal{EF}(\Omega, \mathbb{R})$ sei die Normalform von φ. Dann gilt die Abschätzung $|\varphi|^p \leq \sum_{j=1}^m |\alpha_j|^p \; 1_{A_j}$, also $\|\varphi\|_p < \infty$. Mit Satz 7.3 folgt die Behauptung. $\square$

Das $\mathcal{N}$ aus Bemerkung 7.8 ist für jedes $p \in \{0\} \cup [1, \infty]$ ein Untervektorraum von $\mathcal{L}^p$. Dabei folgen wir für $p = 0$ der Benennung $\mathcal{L}^0 := \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{R})$. Für $p = 0$ ist dies klar, also ist $\mathcal{N}$ ein Vektorraum. Für $p \in [1, \infty]$ folgt dies aus Bemerkung 7.8 (c) $\Rightarrow$ (a).

Die Quotientenräume

$$L^p(\Omega, \mathcal{A}, \mu) := \mathcal{L}^p(\Omega, \mathcal{A}, \mu)/\mathcal{N}, \;\; p \in \{0\} \cup [1, \infty],$$

sind somit wohldefinierte $\mathbb{R}$-Vektorräume. Es gilt

$$L^p(\Omega, \mathcal{A}, \mu) \subset L^0(\Omega, \mathcal{A}, \mu), \;\; p \in [1, \infty].$$

Es bezeichne $[f] \in L^0(\Omega, \mathcal{A}, \mu)$ und g einen Repräsentanten von $[f]$, dann ist $f - g \in \mathcal{N}$, also stimmen f und g μ–f.ü. überein. Mit Bemerkung 7.8 ist $\|[f]\|_p := \|f\|_p$ für jedes $p \in [1, \infty]$ und $f \in L^0(\Omega, \mathcal{A}, \mu)$ wohldefiniert. Für $[f] \in L^p(\Omega, \mathcal{A}, \mu)$ gilt

$$\|[f]\|_p = 0 \Leftrightarrow f = 0 \;\; \mu\text{–f.ü.} \Leftrightarrow [f] = 0.$$

Also ist hier $\|\cdot\|_p$ eine *Norm* auf $L^p(\Omega, \mathcal{A}, \mu)$. Der Preis für diese Verbesserung ist die Tatsache, dass die Elemente von L^p keine Funktionen über Ω, sondern Nebenklassen bezüglich des Untervektorraumes $\mathcal{N}$ von $\mathcal{L}^p$ sind.

Die folgende Vereinbarung führt aber in der Regel zu keinen Mißverständnissen: Für die Nebenklassen $[f] = f + \mathcal{N}$ aus L^p, $p \in \{0\} \cup [1, \infty]$, schreiben wir wieder f und fortan sei

$$L^p := L^p(\Omega, \mathcal{A}, \mu) := (L^p(\Omega, \mathcal{A}, \mu), \|\cdot\|_p), \;\; p \in [1, \infty].$$

Wir fassen zusammen:

7.10 Satz *Für $p \in [1, \infty]$ ist L^p bzgl. $\|\cdot\|_p$ ein normierter $\mathbb{R}$-Vektorraum.*

7.11 Definition Es seien $p \in [1, \infty]$ und $f_n \in \mathcal{L}^p$, $n \in \mathbb{N}$. Die Folge $(f_n)_n$ heißt im *p-ten Mittel konvergent* gegen $f \in \mathcal{L}^p$, falls $\lim_{n\to\infty} \|f_n - f\|_p = 0$. Die Folge $(f_n)_n$ heißt eine CAUCHY-*Folge in $\mathcal{L}^p$* oder eine CAUCHY-Folge für die Konvergenz im p-ten Mittel, falls zu jedem $\epsilon > 0$ ein $N_0(\epsilon) \in \mathbb{N}$ existiert mit $\|f_m - f_n\|_p < \epsilon$ für alle $m, n \geq N_0(\epsilon)$. Analoge Begriffe führt man für L^p statt $\mathcal{L}^p$ ein.

Eine positive Antwort auf die Frage nach der Vollständigkeit von $\mathcal{L}^p$ bzw. L^p gibt der folgende Satz:

7.12 Satz (von Riesz-Fischer) *Für $p \in [1, \infty]$ ist $\mathcal{L}^p(\Omega, \mu, \mathbb{R})$ vollständig.*

7.13 Korollar *Für $p \in [1, \infty]$ ist $L^p(\Omega, \mu, \mathbb{R})$ ein Banachraum.*

Wir geben den Beweis von Satz 7.12 hier an, besprechen ihn in der Vorlesung jedoch nicht.

Beweis von Satz 7.12: Es seien $p \in [1, \infty]$ und $(f_n)_n$ eine CAUCHY-Folge in $\mathcal{L}^p$. Es existiert zu $k \in \mathbb{N}$ ein $N(k)$ mit $\|f_n - f_m\|_p \leq \frac{1}{2^k}$ für $n, m \geq N(k)$. Ohne Einschränkung der Allgemeinheit ist $N(1) < N(2) < N(3) < \cdots$. Wir betrachten die Funktionenfolge

$$g_k := |f_{N(1)}| + \sum_{j=1}^{k} |f_{N(j+1)} - f_{N(j)}|, \quad k \in \mathbb{N}.$$

Jedes g_k ist $\mathcal{A}$-messbar, nicht-negativ und $g_k \leq g_{k+1}$.

Es folgt mit der MINKOWSKI-Ungleichung, Satz 7.6,

$$\|g_k\|_p \leq \|f_{N(1)}\|_p + \sum_{j=1}^{k} \|f_{N(j+1)} - f_{N(j)}\|_p \leq \|f_{N(1)}\|_p + \sum_{j=1}^{k} \frac{1}{2^j} \leq \|f_{N(1)}\|_p + 1 < \infty\,.$$

Mit dem Satz über die monotone Konvergenz (Satz 5.10) folgt

$$\lim_{k\to\infty} \|g_k\|_p^p = \lim_{k\to\infty} \int_\Omega g_k^p \, d\mu = \int_\Omega \big(\lim_{k\to\infty} g_k^p\big) \, d\mu < \infty.$$

Also ist $\lim_{k\to\infty} g_k^p$ μ–f.ü. endlich (Korollar 6.4). Es existiert also eine μ-Nullmenge $A \in \mathcal{A}$ mit

$$\lim_{k\to\infty} g_k(x) < \infty$$

für $x \in A^c$. Zu $x \in A^c$ ist

$$\lim_{k\to\infty} g_k(x) = |f_{N(1)}(x)| + \sum_{j=1}^{\infty} |f_{N(j+1)}(x) - f_{N(j)}(x)|.$$

Also ist $f_{N(1)}(x) + \sum_{j\geq 1}\big(f_{N(j+1)}(x) - f_{N(j)}(x)\big)$ in $\mathbb{R}$ absolut konvergent, also insbesondere konvergent, und die $(k-1)$-te Partialsumme ist $f_{N(k)}(x)$. Also existiert $\lim_{k\to\infty} f_{N(k)}(x)$. Man setzt nun

$$f(x) := \begin{cases} \lim_{k\to\infty} f_{N(k)}(x)\ , & x \in A^c, \\ 0\ , & x \in A. \end{cases}$$

Dann ist $f|_{A^c}$ $\mathcal{A}$-messbar, und $f|_A$ ist es auch, also ist f $\mathcal{A}$-messbar.

Es bleibt zu zeigen, dass $(f_n)_n$ im p-ten Mittel gegen f konvergiert. Zu $\epsilon > 0$ existiert ein $N(\epsilon) \in \mathbb{N}$ mit $\|f_n - f_m\|_p < \epsilon$ für $n, m \geq N(\epsilon)$. Aus dem Lemma von FATOU (Satz 5.14) folgt

$$\begin{aligned} \|f - f_n\|_p^p &= \int_\Omega \big|\lim_{k\to\infty} f_{N(k)}(x) - f_n(x)\big|^p\, d\mu(x) = \int_\Omega \lim_{k\to\infty} \big|f_{N(k)}(x) - f_n(x)\big|^p\, d\mu(x) \\ &\leq \liminf_{k\to\infty} \int_\Omega \big|f_{N(k)}(x) - f_n(x)\big|^p\, d\mu(x) = \liminf_{k\to\infty} \|f_{N(k)} - f_n\|_p^p \leq \varepsilon^p \end{aligned}$$

für $n > N(\epsilon)$. Weiterhin impliziert

$$\|f\|_p \leq \|f - f_n\|_p + \|f_n\|_p < \epsilon + \|f_n\|_p < \infty,$$

dass $f \in \mathcal{L}^p$ und die obige Abschätzung liefert die gewünschte Konvergenz.

Es seien nun $p = \infty$ und $(f_n)_n$ eine CAUCHY-Folge in $\mathcal{L}^\infty$. Wir setzen

$$N := \bigcup_{n\geq 1} \big[|f_n| > \|f_n\|_\infty\big] \cup \bigcup_{m,n\geq 1} \big[|f_m - f_n| > \|f_m - f_n\|_\infty\big].$$

Dies ist eine μ-Nullmenge und für $x \in N^c$ ist

$$|f_m(x) - f_n(x)| \leq \|f_m - f_n\|_\infty,\ \ m, n \in \mathbb{N}.$$

Also konvergiert $(f_n)_n$ auf N^c gleichmäßig gegen $f := \lim_{n\to\infty} 1_{N^c} f_n$. Insbesondere ist $f \in \mathcal{L}^\infty$ und $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$. Damit ist der Satz bewiesen. □

Dem Beweis des Satzes von RIESZ-FISCHER entnimmt man das folgende Resultat:

7.14 Korollar (von H. Weyl) *Es sei $p \in [1, \infty]$.*

(a) *Zu jeder* CAUCHY*-Folge* $(f_n)_n$ *in* $\mathcal{L}^p$ *gibt es eine Teilfolge* $(f_{n_k})_k$ *und ein* $f \in \mathcal{L}^p$*, so dass* $f_{n_k} \to f$ μ*–f.ü.*

(b) *Konvergiert die Folge* $(f_n)_n$ *in* $\mathcal{L}^p$ *im* p*-ten Mittel gegen* $f \in \mathcal{L}^p$*, so existiert eine Teilfolge* $(f_{n_k})_k$*, die* μ*–f.ü. gegen* f *konvergiert.*

Beweis: zu (a): Siehe Beweis des Satzes von RIESZ-FISCHER.

zu (b): Nach Beweis des Satzes von RIESZ-FISCHER gibt es ein $g \in \mathcal{L}^p$ mit $\|f_n - g\|_p \to 0$ und eine Teilfolge $(f_{n_k})_k$, die μ–f.ü. gegen g konvergiert. Wegen $\|f_n - f\|_p \to 0$ ist dann $f = g$ μ–f.ü. □

7.15 Definition Sei X eine P-integrierbare Zufallsgröße auf $(\Omega, \mathcal{A}, P)$ und $\mu := \mathbb{E}(X)$, so heißt

$$\operatorname{Var} X := \mathbb{E}(X-\mu)^2 = \int (X-\mu)^2 \, dP = \int (x-\mu)^2 \, dP^X(x)$$

Varianz von X.

$$\sigma(X) := (\operatorname{Var} X)^{1/2} = (\mathbb{E}(X-\mu)^2)^{1/2}$$

heißt *Standardabweichung* von X. Für $k \in \mathbb{N}$ nennt man $\mathbb{E}X^k$ und $\mathbb{E}(X-\mu)^k$, wenn diese Größen existieren, das *k-te Moment* bzw. das *zentrale k-te Moment*, sowie für $p > 0$ nennt man $\mathbb{E}|X|^p$ und $\mathbb{E}|X-\mu|^p$ das *p-te absolute* bzw. das *zentrale p-te absolute Moment*.

7.16 Satz *Eine Zufallsgröße* X *auf einem W-Raum* $(\Omega, \mathcal{A}, P)$ *ist genau dann quadratisch integrierbar, wenn* X *integrierbar und* $\operatorname{Var} X < \infty$ *ist. Es gilt dann:*

$$\begin{aligned} \operatorname{Var} X &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \int x^2 \, dP^X(x) - \Big(\int x \, dP^X(x) \Big)^2 . \end{aligned}$$

Für integrierbares X *gilt stets*

$$\begin{aligned} \mathbb{E}(X)^2 &\le \mathbb{E}(X^2) \quad \textit{sowie} \\ \operatorname{Var}(\alpha X + \beta) &= \alpha^2 \operatorname{Var} X, \quad \alpha, \beta \in \mathbb{R}. \end{aligned}$$

X *hat genau dann Varianz* 0*, wenn* X P*-f.s. konstant ist.*

Für P-fast überall sagt man in der Wahrscheinlichkeitstheorie auch P-*fast sicher*, abgekürzt P-f.s.

Beweis: Es gilt $\mathcal{L}^2(P) \subset \mathcal{L}^1(P)$ mittels der HÖLDER-Ungleichung. Alle konstanten reellen Funktionen sind in $\mathcal{L}^2(P)$. Daher folgt aus $X \in \mathcal{L}^2(P)$ die Integrierbarkeit sowie $\operatorname{Var} X < \infty$. Sei umgekehrt $X \in \mathcal{L}^1(P)$ und $\operatorname{Var} X < \infty$, so liegt $X - \mathbb{E}(X) \in \mathcal{L}^2(P)$, also auch $X = X - \mathbb{E}(X) + \mathbb{E}(X)$. Da $\mathbb{E}$ linear ist, folgt $\operatorname{Var} X = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. $\mathbb{E}(X^2) \le \mathbb{E}(X)^2$ ist dann klar. Es gilt

$$\begin{aligned} \operatorname{Var}(\alpha X + \beta) &= \mathbb{E}((\alpha X + \beta) - (\alpha \mathbb{E} X + \beta))^2 \\ &= \mathbb{E}(\alpha X - \alpha \mathbb{E} X)^2 = \alpha^2 \operatorname{Var} X \ . \end{aligned}$$

Hat X Varianz 0, also $\mathbb{E}(X - \mathbb{E}X)^2 = 0$, so ist dies äquivalent zu $(X - \mathbb{E}X)^2 = 0$ P-f.s., das heißt $X = \mathbb{E}X$ P-f.s. □

Wir kommen zu einer Reihe von weiteren wichtigen Ungleichungen.

7.17 Satz *X, Y seien Zufallsgrößen auf einem W-Raum $(\Omega, \mathcal{A}, P)$. Dann gilt:*

(a) (MARKOV-Ungleichung) *Für jedes $t \ge 0$ und jede monoton wachsende Funktion $g : [0, \infty) \to [0, \infty)$ mit $g(t) > 0$ ist*

$$P(|X| \ge t) \le g(t)^{-1} \int_{\{|X| \ge t\}} g(|X|) \, dP \le g(t)^{-1} \mathbb{E} g(|X|)$$

(b) (TSCHEBYSCHEV-Ungleichung) *Speziell im Fall $g(t) = t^2$ und $\mathbb{E}X = 0$ folgt*

$$P(|X| \ge t) \le t^{-2} \int_{\{|X| \ge t\}} X^2 \, dP \le \frac{\operatorname{Var} X}{t^2}$$

(c) (JENSENsche Ungleichung) *Es sei $I \subset \mathbb{R}$ ein Intervall, $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit $P(X \in I) = 1$, P-integrierbar und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion, dann ist $\int X \, dP = \mathbb{E}X \in I$, $\varphi \circ X$ ist quasi-integrierbar und es gilt*

$$\varphi(\mathbb{E}X) = \varphi\left(\int_\Omega X \, dP \right) \le \int_\Omega \varphi \circ X \, dP = \mathbb{E}(\varphi(X))$$

Beweis: zu (a): Verwende $g(t) 1_{\{|X| \ge t\}} \le g(|X|) 1_{\{|X| \ge t\}} \le g(|X|)$ und die Monotonie des Integrals, (b) ist klar.

zu(c): φ heißt konvex, wenn für alle $x, y \in I$ und $\lambda \in [0, 1]$ gilt:

$$\varphi(\lambda x + (1 - \lambda) y) \le \lambda \varphi(x) + (1 - \lambda) \varphi(y) \ .$$

In Analysis I zeigt man, dass dazu äquivalent ist

$$\frac{\varphi(t) - \varphi(x)}{t - x} \le \frac{\varphi(y) - \varphi(x)}{y - x} \quad \forall x, y, t \in I : x < t < y$$

bzw.

$$\frac{\varphi(t)-\varphi(x)}{t-x} \leq \frac{\varphi(y)-\varphi(t)}{y-t} \quad \forall x,y,t \in I : x < t < y.$$

Daraus folgt, dass φ auf $\mathring{I}$ stetig ist, denn sei $x_0 \in \mathring{I}$, $s,t \in \mathring{I}$ mit $s < x_0 < t$. Ist nun $x_0 < x < t$, so folgt aus obigen Ungleichungen

$$\frac{\varphi(x_0)-\varphi(s)}{x_0-s} \leq \frac{\varphi(x)-\varphi(x_0)}{x-x_0} \leq \frac{\varphi(t)-\varphi(x_0)}{t-x_0} \ .$$

Daraus folgt sofort die rechtsseitige Stetigkeit. Die linksseitige Stetigkeit beweist man analog. Also ist φ höchstens in den Randpunkten von I unstetig und somit ist jede konvexe Funktion BOREL-messbar! Zum Beweis der Ungleichung: Wir zeigen zunächst $m := \mathbb{E}(X) \in I$. $a, b \in \bar{\mathbb{R}}$ seien linker bzw. rechter Randpunkt von I. Mit $a \leq X \leq b$ folgt $a \leq \mathbb{E}(X) \leq b$. Ist nun $a \in \mathbb{R}$ und $a \notin I$, so ist $0 < X(\omega) - a$ für alle $\omega \in \Omega$, also $a < m$ nach Satz 6.3(a). Analoges folgt für b, also $m \in I$.

Ist m kein innerer Punkt von I, so ist $m \in \mathbb{R}$ rechter oder linker Randpunkt von I. Also ist $X(\omega) = m$ für P-fast alle $\omega \in \Omega$, also $\varphi(X(\omega)) = \varphi(\mathbb{E}X) = \varphi(m)$ für P-fast alle $\omega \in \Omega$, also $\mathbb{E}(\varphi X) = \varphi(\mathbb{E}X)$.

Es sei nun $m \in \mathring{I}$. Nun konstruieren wir eine Stützgerade an den Graphen von φ im Punkt $(m, \varphi(m))$: Für $s,t \in I$ mit $s < m < t$ ist $\frac{\varphi(m)-\varphi(s)}{m-s} \leq \frac{\varphi(t)-\varphi(m)}{t-m}$, also ist $\alpha := \sup\{\frac{\varphi(m)-\varphi(s)}{m-s},\ s < m,\ s \in I\} < \infty$ und für alle $t \in I$ mit $t > m$ gilt:

$$\varphi(t) \geq \varphi(m) + \alpha(t-m) \tag{7.1}$$

Für $t = m$ ist (7.1) auch richtig und sie gilt nach Definition von α auch für alle $t \in I$ mit $t < m$. Somit gilt (7.1) für alle $t \in I$. Der Graph von φ auf I verläuft also stets oberhalb der durch $t \mapsto \varphi(m) + \alpha(t-m)$ definierten Stützgeraden. Es folgt

$$\varphi(X(\omega)) \geq \varphi(\mathbb{E}X) + \alpha(X(\omega) - \mathbb{E}X)$$

Integration dieser Ungleichung nach P liefert die JENSENsche Ungleichung. □

7.18 Korollar *Es sei $X \in \mathcal{L}^s(\Omega, \mathcal{A}, P)$. Dann ist $X \in \mathcal{L}^r$ für $1 \leq r \leq s$. Ist X P-f.s. beschränkt, $X \in \mathcal{L}^\infty(\Omega, \mathcal{A}, P)$, so gilt*

$$||X||_r \uparrow ||X||_\infty \quad (r \to \infty).$$

Beweis: Es sei $\varphi(t) = t^{s/r}$, $t \geq 0$. Für $r \in [1, s]$ folgt aus der JENSENschen Ungleichung:

$$(\mathbb{E}|X|^r)^{1/r} = \Big(\varphi(\mathbb{E}|X|^r)\Big)^{1/s} \leq \Big(\mathbb{E}\varphi(|X|^r)\Big)^{1/s} = \mathbb{E}(|X|^s)^{1/s} \ . \tag{7.2}$$

Ist X P-f.s. beschränkt, so ist X r-fach integrierbar für jedes $r \geq 1$. Mit (7.2) folgt die Konvergenz von $||X||_r$ gegen einen Limes a. Aus $|X| \leq ||X||_\infty$ P-f.s. folgt $\|X\|_r \leq ||X||_\infty$ für jedes $r \geq 1$, also $a \leq ||X||_\infty$. Nun zeigen wir noch $a \geq ||X||_\infty$: Für $0 < c < ||X||_\infty$ ist $P(|X| > c) > 0$. Weiter gilt

$$\begin{aligned} (\mathbb{E}|X|^r)^{1/r} &\geq \left(\int_{\{|X|>c\}} |X|^r \, dP \right)^{1/r} \\ &\geq cP(|X| > c)^{1/r}. \end{aligned}$$

Also ist $a = \lim_{r\to\infty}(\mathbb{E}|X|^r)^{1/r} \geq c$ für alle $c < ||X||_\infty$, also ist $a \geq ||X||_\infty$. $\square$

7.19 Definition Ist $X = (X_1, \ldots, X_n)$ ein Zufallsvektor, so definiert man den Erwartungswert komponentenweise durch $\mathbb{E}(X) = (\mathbb{E}X_1, \ldots, \mathbb{E}X_n) \in \mathbb{R}^n$.

7.20 Definition Sind X und Y aus $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ mit $X \cdot Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, so ist ihre *Kovarianz* $\mathrm{Cov}(X, Y)$ definiert durch

$$\begin{aligned} \mathrm{Cov}(X, Y) &:= \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \; . \end{aligned}$$

Ist $X = (X_1, \ldots, X_n)$ ein Zufallsvektor mit $X_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, $i = 1, \ldots, n$, und $X_iX_j \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ für alle $i, j \in \{1, \ldots, n\}$, so ist die *Kovarianzmatrix* $\Sigma(X) = (\sigma_{ij}(X))$ definiert durch $\sigma_{ij}(X) = \mathrm{Cov}(X_i, X_j)$.

Offenbar ist $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$ für eine eindimensionale Zufallsgröße X. Ist X ein Zufallsvektor, als Spaltenvektor geschrieben, so ist

$$\Sigma(X) = \mathbb{E}((X - \mathbb{E}X)(X - \mathbb{E}X)^T).$$

7.21 Satz

(a) *Sind* $X, Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$, *so existiert* $\mathrm{Cov}(X, Y)$ *und*

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$

(b) *Für* $X_1, \ldots, X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ *gilt*

$$\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \sum_{1\leq i<j\leq n} \mathrm{Cov}(X_i, X_j).$$

(c) *Die Kovarianzmatrix ist symmetrisch und positiv semidefinit.*

Beweis: (a) folgt aus CAUCHY-SCHWARZ und einfachem Nachrechnen und (b) analog. (c): Für $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ gilt

$$0 \leq \mathbb{E}\Big(\sum_{i=1}^n \lambda_i (X_i - \mathbb{E}X_i) \Big)^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \operatorname{Cov}(X_i, X_j) \ .$$

□

7.22 Definition Zwei quadratisch integrierbare Zufallsgrößen X und Y heißen *unkorreliert*, wenn ihre Kovarianz verschwindet, d.h.

$$\operatorname{Cov}(X, Y) = 0.$$

Für eine Menge $X_1, \dots, X_n$ von Zufallsgrößen mit endlicher Varianz gilt

$$\operatorname{Var}(\sum_{j=1}^n X_j) = \sum_{j=1}^n \operatorname{Var}(X_j),$$

wenn $X_1, \dots, X_n$ *paarweise unkorreliert* sind.

7.23 Satz $X, Y, Z \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$*;* $\alpha, \beta \in \mathbb{R}$*. Es gilt*

(a) $\operatorname{Cov}(\alpha X + \beta Y, Z) = \alpha \operatorname{Cov}(X, Z) + \beta \operatorname{Cov}(Y, Z)$

(b) $\operatorname{Cov}(X, \alpha Y + \beta Z) = \alpha \operatorname{Cov}(X, Y) + \beta \operatorname{Cov}(X, Z)$

(c) $|\operatorname{Cov}(X, Y)| \leq (\operatorname{Var}(X))^{1/2} (\operatorname{Var}(Y))^{1/2}$

Insgesamt ist also $\operatorname{Cov}(\cdot, \cdot)$ *eine symmetrische Bilinearform auf* $\mathcal{L}^2$*.*

Beweis: Nachrechnen und bei (c) CAUCHY-SCHWARZ. □

7.24 Satz *Sei* $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ *eine Zufallsgröße. Dann gilt*

$$\operatorname{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \min_{a \in \mathbb{R}} \mathbb{E}(X - a)^2$$

Beweis: Für $a \in \mathbb{R}$ ist $\mathbb{E}(X-a)^2 = \mathbb{E}(X - \mathbb{E}X)^2 + 2\mathbb{E}(X-a)\mathbb{E}(X - \mathbb{E}X) + (\mathbb{E}X - a)^2$. Da $\mathbb{E}(X - \mathbb{E}X) = 0$, wird $\mathbb{E}(X-a)^2$ für $a = \mathbb{E}(X)$ minimiert.

Wir wollen uns speziellen Situationen und Beispielen zu den neuen Begriffen in diesem Kapitel widmen:

7.25 Beispiel Sei $X \geq 0$ eine Zufallsgröße auf $(\Omega, \mathcal{A}, P)$ und $X(\Omega)$ abzählbar. Dann ist

$$\mathbb{E}(X) = \mathbb{E}\Big(\sum_{x \in X(\Omega)} x 1_{\{X = x\}} \Big) = \sum_{x \in X(\Omega)} x P(X = x) \ .$$

Falls nicht notwendig $X \geq 0$, jedoch X quasi-integrierbar ist, so gilt

$$\mathbb{E}(X) = \sum_{\substack{x \in X(\Omega) \\ x \geq 0}} x P(X = x) - \sum_{\substack{x \in X(\Omega) \\ x < 0}} (-x) P(X = x).$$

Ist insbesondere Ω abzählbar und $X \geq 0$, so gilt $X = \sum_{\omega \in \Omega} X(\omega) 1_{\{\omega\}}$, also

$$\begin{aligned} \mathbb{E}(X) &= \sum_{\omega \in \Omega} X(\omega) \mathbb{E}(1_{\{\omega\}}) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \\ &= \sum_{\omega \in \Omega} X(\omega) p(\omega), \quad \text{mit } \; p(\omega) = P(\{\omega\}). \end{aligned}$$

7.26 Beispiel (*Fairer Münzwurf*) $\Omega = \{0,1\}^{\mathbb{N}}$, und $\mathcal{A}$, P wie in Beispiel 1.13(c).

(a) Sei $X_j\Big((x_i)_i\Big) := x_j$ für $(x_i)_i \in \Omega$ und $j \in \mathbb{N}$. Dann ist

$$\mathbb{E}(X_j) = 1 \cdot P(X_j = 1) + 0 \cdot P(X_j = 0) = 1/2 \; .$$

(b) Sei $S_n := X_1 + \cdots + X_n$ (= Anzahl der Erfolge in n Würfen) mit $(X_j)_{j=1,\dots,n}$ wie in (a), dann ist für $k \in \{0, \dots, n\}$

$$P(S_n = k) = \sum_{\substack{(x_1,\dots,x_n) \in \{0,1\}^n \\ \sum_{j=1}^n x_j = k}} P(X_1 = x_1, \dots, X_n = x_n) = \binom{n}{k} 2^{-n}$$

und $\mathbb{E}(S_n) = \sum_{j=1}^n \mathbb{E}(X_j) = n \cdot \frac{1}{2}$.

(c) Sei $T : \Omega \to \mathbb{N} \cup \{\infty\}$ die Wartezeit auf die erste Eins, d.h. $T(\omega) := \min\{n \in \mathbb{N} | X_n(\omega) = 1\}$, dann ist

$$P(T = k) = P(X_1 = X_2 = \cdots = X_{k-1} = 0, X_k = 1) = 2^{-k} \; , \; k \in \mathbb{N},$$

und $\mathbb{E}(T) = \sum_{k=1}^\infty k P(T = k) = \sum_{k=1}^\infty k 2^{-k} = 2$.

(d) Wir starten mit einem Einsatz von einer Einheit und verdoppeln den Einsatz bis 1 auftritt. Der Einsatz in der n-ten Runde ist dann $X_n = 2^{n-1} 1_{\{T > n-1\}}$ mit T wie in (c). Daher ist

$$\mathbb{E}(X_n) = 2^{n-1} P(\{T > n-1\}) = 2^{n-1} \Big(\frac{1}{2}\Big)^{n-1} = 1 \; .$$

Aber $\lim_{n \to \infty} X_n = 0$ P-fast sicher (für alle $\omega \neq (0,0,0,\dots)$).

7.27 Beispiel (*geometrische Wahrscheinlichkeit*) Konstruktion einer Gleichverteilung auf S^1 (Sphäre)

Problem: $S^1 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ als Teilmenge von $\mathbb{R}^2$ hat Lebesgue-Maß 0! Die Abbildung $\Gamma : [0,2\pi) \to S^1$, $\phi \mapsto (\cos\phi, \sin\phi)$ ist bijektiv und $\mathcal{B}^1_{[0,2\pi)}/\mathcal{B}^2_{S^1}$-messbar. Γ überführt jedes Teilintervall I von $[0,2\pi)$ in ein Kreisrandsegment $K_I := \{(\cos\phi, \sin\phi), \phi \in I\}$ derselben Länge (zeigen wir gleich). Sei nun Φ eine auf $[0,2\pi)$ gleichverteilte Zufallsgröße. (Allgemein: Ist Ω eine Borel-messbare Teilmenge des $\mathbb{R}^d$ mit $0 < \lambda^d(\Omega) < \infty$, dann heißt das W-Maß $P : \mathcal{B}^d_\Omega \to [0,1]$, definiert durch $P(A) = \frac{\lambda^d(A)}{\lambda^d(\Omega)}$ *Gleichverteilung auf* Ω.) Dann besitzt $\Gamma \circ \Phi$ eine Verteilung Q mit

$$Q(K_I) = P(\Gamma \circ \Phi \in K_I) = \frac{\lambda(I)}{2\pi} = \frac{\text{Länge von } K_I}{2\pi}$$

für alle Intervalle $I \subset [0,2\pi)$. Q heißt Gleichverteilung auf S^1. Dies sieht man so: Die Länge $L(K)$ einer differenzierbaren Kurve $K : I \to \mathbb{R}^2$ berechnet sich zu $L(K) = \int_a^b ||K'(t)||\,dt$, wobei $||\cdot||$ die euklidische Norm ist und a, b die Randpunkte von I (siehe Analysis III). Für $K_I = \Gamma(I)$ liefert das speziell:

$$L(K_I) = \int_a^b ||\Gamma'(\phi)||\,d\phi = b - a = \lambda(I),$$

denn $\Gamma'(\phi) = (-\sin(\phi), \cos(\phi))$ und somit $||\Gamma'(\phi)|| = 1$ für alle $\phi \in [0,2\pi)$. Da $P(\Gamma \circ \Phi \in K_I) = P(\Phi \in I) = \frac{\lambda(I)}{2\pi}$, folgt die Behauptung.

7.28 Beispiele

(a) *Binomialverteilung*

Es sei X nach $b(n,p)$ verteilt (also $X = X_1 + \cdots + X_n$ mit X_i, so dass $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$, $0 \le p \le 1$, wissen wir aber noch gar nicht!)

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^n k\binom{n}{k}p^k(1-p)^{n-k} = np(p + (1-p))^{n-1} = np \\
\mathbb{E}(X^2) &= \sum_{k=1}^n k^2\binom{n}{k}p^k(1-p)^{n-k} = np\sum_{k=1}^n k\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} \\
&= np[(n-1)p + 1] = np(np + (1-p)),
\end{aligned}$$

also $\mathrm{Var}(X) = np(1-p)$.
Aber: Bitte, diese Herleitung nicht merken.
Sei $X' = 2X - n$. Die Verteilung von X' ist das Bild der Verteilung von X unter der Abbildung $x \mapsto 2x - n$. Also besitzt X' die Verteilung

$$\sum_{k=0}^n \binom{n}{k}p^k(1-p)^{n-k}\delta_{2k-n} =: b_s(n,p)\ .$$

X' hat die Werte $-n, -n+2, \ldots, n-2, n$ und heißt daher *symmetrische* oder *symmetrisierte Binomialverteilung*. Speziell ist $b_s(1,p) = (1-p)\delta_{-1} + p\delta_1$.

(b) POISSON-*Verteilung* X sei π_α-verteilt, dann ist

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{\infty} e^{-\alpha}\frac{\alpha^k}{k!}k = \alpha \quad \text{und} \\ \mathbb{E}(X^2) &= \sum_{k=0}^{\infty} e^{-\alpha}\frac{\alpha^k}{k!}k^2 = \sum_{k=1}^{n} e^{-\alpha}\frac{\alpha^k}{(k-1)!}(k-1+1) = \alpha^2 + \alpha,\end{aligned}$$

also $\operatorname{Var}(X) = \alpha$.

(c) *Normalverteilung* Sei $\nu_{\alpha,\sigma^2} := g_{\alpha,\sigma^2} \cdot \lambda^1$, so gilt $T(\nu_{0,1}) = \nu_{\alpha,\sigma^2}$ für $T(x) = \sigma x + \alpha$, $\alpha \in \mathbb{R}$, $\sigma > 0$. Jede $N(0,1)$–verteilte Zufallsvariable X ist p–fach integrierbar für jedes $p \geq 0$, denn aus $\frac{t^k}{k!} < e^t$ für $t > 0$, $k = 0, 1, 2, \ldots$ folgt

$$|x|^p e^{-x^2/2} < 2^k k! |x|^{p-2k} \quad (x \neq 0) \ .$$

Nun ist $x \mapsto |x|^\alpha$ für $\alpha < -1$ über $\mathbb{R} \setminus [-1,1]$ λ^1–integrierbar und somit ist $x \mapsto |x|^p\, g_{0,1}(x)$ offenbar λ^1–integrierbar über $\mathbb{R}$ für jedes $p \geq 0$, indem man k hinreichend groß wählt. Also existieren

$$M_n := \mathbb{E}(X^n) = \int_{-\infty}^{\infty} x^n g_{0,1}(x)\, dx$$

für ganzzahlige $n \geq 0$. Natürlich ist $M_{2k-1} = 0$, $k \in \mathbb{N}$, da hier der Integrand ungerade ist. Also $\mathbb{E}(X) = 0$ für $N(0,1)$–verteiltes X und $\mathbb{E}(X) = \alpha$ für $N(\alpha,\sigma^2)$–verteiltes X. Für gerades $n \geq 2$ gilt

$$\int_{-\infty}^{\infty} x^n g_{0,1}(x)\, dx = -x^{n-1} g_{0,1}(x)\big|_{-\infty}^{+\infty} + (n-1)\int_{-\infty}^{\infty} x^{n-2} g_{0,1}(x)\, dx \ ,$$

denn $g_{0,1}$ ist Stammfunktion von $x \mapsto -x g_{0,1}(x)$ und alle Integranden sind stetig, so dass diese Integrale auch als absolut konvergente RIEMANN–Integrale existieren. Wir können also partielle Integration anwenden. Es folgt

$$M_{2k} = (2k-1) M_{2k-2} \ , \ k \in \mathbb{N} \ .$$

Mit $M_0 = 1$ folgt $M_{2k} = 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2k-1)$, $k \in \mathbb{N}$. Insbesondere ist $\mathbb{E}(X^2) = 1$ für $N(0,1)$–verteiltes X und $\mathbb{E}(X^2) = \alpha^2 + \sigma^2$ für $N(\alpha,\sigma^2)$–verteiltes X. Somit ist $\operatorname{Var}(X) = \sigma^2$. Weiter folgt für $N(\alpha,\sigma^2)$–verteiltes X:

$$\mathbb{E}(X^n) = \sum_{k=0}^{n} \binom{n}{k} \sigma^k \alpha^{n-k} M_k \ .$$

Dazu verwende $T(x) = \sigma x + \alpha$ und

$$\begin{aligned}
\mathbb{E}(X^n) &= \int_{\mathbb{R}} x^n \, d\nu_{\alpha,\sigma^2}(x) = \int_{\mathbb{R}} x^n \, dT(\nu_{0,1})(x) \\
&= \int_{\mathbb{R}} (\sigma x + \alpha)^n \, d\nu_{0,1}(x) \ .
\end{aligned}$$

(d) CAUCHY-*Verteilung* γ_α: X sei CAUCHY–verteilt mit Parameter $\alpha > 0$. Der Erwartungswert existiert nicht, denn aus $\int_0^n \frac{x}{1+x^2} \, dx = \frac{1}{2} \log(1+n^2)$ folgt

$$\mathbb{E}(X^+) = \mathbb{E}(X^-) = \int_0^\infty \frac{x}{1+x^2} \, dx = +\infty \ .$$

X ist nicht einmal quasi–integrierbar.

(e) *Mehrdimensionale Normalverteilung* auf $\mathbb{R}^n$: X sei standardnormalverteilt auf $\mathbb{R}^n$, dann ist

$$\mathbb{E}(X_i) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i \exp(-\frac{1}{2} \sum_{k=1}^n x_k^2) \, d\lambda^n(x) = 0 \ .$$

Für $i, j \in \{1, \dots, n\}$ mit $i \neq j$ gilt

$$\mathbb{E}(X_i X_j) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} x_i x_j \exp(-\frac{1}{2} \sum_{k=1}^n x_k^2) \, d\lambda^n(x) = 0$$

und weiter gilt

$$\mathbb{E}(X_i^2) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} x_i^2 \exp(-\frac{1}{2} \sum_{k=1}^n x_k^2) \, d\lambda^n(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x_i^2 e^{-x_i^2/2} \, d\lambda(x_i) = 1 \ .$$

Also ist $\Sigma(X)$ die Einheitsmatrix.

Sei X ein n–dimensionaler Zufallsvektor mit Kovarianzmatrix $\Sigma(X)$ und Erwartungswert $a \in \mathbb{R}^n$. Weiter sei A eine $m \times n$–Matrix, $b \in \mathbb{R}^m$ und $Y := AX + b$, dann ist $\mathbb{E}(Y) = Aa + b$ und

$$\begin{aligned}
\Sigma(Y) &= \mathbb{E}\Big((Y - \mathbb{E}(Y))(Y - \mathbb{E}(Y))^t\Big) \\
&= \mathbb{E}\Big(A(X-a)(X-a)^t A^t\Big) = A\Sigma(X)A^t \ .
\end{aligned}$$

Also ist für die allgemeine Normalverteilung in 5.30(b) die Kovarianzmatrix gleich AA^t und der Vektor der Erwartungswerte gleich b.

Wir schließen dieses Kapitel mit einer nützlichen Abschätzung.

7.29 Satz *Sei X eine $N(0,\sigma^2)$–verteilte Zufallsgröße, dann gilt für alle $\eta > 0$:*

$$(2\pi)^{-\frac{1}{2}}\,\frac{\sigma\eta}{\sigma^2+\eta^2}\,e^{-\frac{\eta^2}{2\sigma^2}} < P(X \geq \eta) < (2\pi)^{-\frac{1}{2}}\,\frac{\sigma}{\eta}\,e^{-\frac{\eta^2}{2\sigma^2}} \ .$$

Beweis: Wir zeigen den Fall $\sigma = 1$. Der allgemeine Fall folgt, indem man X durch σX ersetzt, und somit η durch η/σ. Es gilt

$$P(X \geq \eta) = P^X([\eta,\infty)) = \nu_{0,1}([\eta,\infty)) = (2\pi)^{-1/2}\int_\eta^\infty e^{-x^2/2}\,dx$$

und

$$\int_\eta^\infty e^{-x^2/2}\,dx < \int_\eta^\infty \frac{x}{\eta}e^{-x^2/2}\,dx = \frac{1}{\eta}\int_\eta^\infty xe^{-x^2/2}\,dx = \frac{1}{\eta}e^{-\eta^2/2} \ .$$

Mittels partieller Integration folgt andererseits

$$\int_\eta^\infty x^{-2}e^{-x^2/2}\,dx = \frac{1}{\eta}e^{-\eta^2/2} - \int_\eta^\infty e^{-x^2/2}\,dx \ ,$$

also

$$\frac{1}{\eta}e^{-\eta^2/2} = \int_\eta^\infty (1+x^{-2})e^{-x^2/2}\,dx < \int_\eta^\infty (1+\eta^{-2})e^{-x^2/2}\,dx \ ,$$

und dazu äquivalent

$$\frac{\eta}{1+\eta^2}e^{-\eta^2/2} < \int_\eta^\infty e^{-x^2/2}\,dx \ .$$

□

Den nächsten Satz werden wir in den folgenden Kapiteln noch deutlich verschärfen.

7.30 Satz (schwaches Gesetz der großen Zahlen) *Für jedes $n \in N$ seien paarweise unkorrelierte Zufallsgrößen $X_1,\dots,X_n$ gegeben, die alle den gleichen Erwartungswert $E \in \mathbb{R}$ und die gleiche Varianz $V < \infty$ besitzen. Sei $\frac{1}{n}S_n := \frac{1}{n}(X_1+\cdots+X_n)$, dann gilt für jedes $\varepsilon > 0$*

$$\lim_{n\to\infty} P\Big(\Big|\frac{1}{n}S_n - E\Big| > \varepsilon\Big) = 0.$$

Beweis: Die Linearität des Erwartungswerts liefert $\mathbb{E}(\frac{1}{n}S_n) = E$ und auf Grund der paarweisen Unkorreliertheit ist

$$\mathrm{Var}\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}\,\mathrm{Var}(X_1+\cdots+X_n) = \frac{1}{n^2}n\,\mathrm{Var}(X_1) = \frac{1}{n}V \ .$$

Also folgt mit der TSCHEBYSCHEV-Ungleichung

$$P\Big(\Big|\frac{1}{n}S_n - E\Big| > \varepsilon\Big) \leq \frac{\mathrm{Var}(\frac{1}{n}S_n)}{\varepsilon^2} \ ,$$

und dies konvergiert gegen Null für $n \to \infty$. $\square$

Man spricht in Satz 7.30 auch von *Konvergenz in Wahrscheinlichkeit* oder *stochastischer Konvergenz*. Dies betrachten wir später genauer.

KAPITEL 8

Produktmaße und der Satz von Fubini

Im Zentrum steht die Konstruktion von Produkträumen und die Diskussion mehrfacher Integrale. Der Satz von FUBINI gestattet die Reduktion mehrfacher Integrale auf einfache: Integrale einer Funktion von mehreren Variablen können iterativ berechnet werden und die Integrationsreihenfolge kann beliebig gewählt werden. In der Wahrscheinlichkeitstheorie möchte man endlich viele oder unendlich viele zufällige Experimente behandeln. Allgemein spricht man von *gekoppelten Zufallsexperimenten*. Die maßtheoretische Grundlage zur mathematischen Beschreibung gekoppelter Experimente ist die Konstruktion eines Produkt-Wahrscheinlichkeitsmaßes auf unendlichen Produkträumen.

Gegeben seien fortan endlich viele Maßräume $(\Omega_j,\, \mathcal{A}_j,\, \mu_j)$, $j = 1, \ldots, d$. Es seien

$$\Omega := \Omega_1 \times \cdots \times \Omega_d = \prod_{j=1}^{d} \Omega_j$$

und $p_j : \Omega \to \Omega_j$, $\omega = (\omega_1, \ldots, \omega_d) \mapsto \omega_j$, $1 \le j \le d$, die *Projektions-Abbildungen.*

8.1 Definition Die von den Projektionen $p_1, \ldots, p_d$ auf $\Omega := \Omega_1 \times \cdots \times \Omega_d$ erzeugte σ-Algebra

$$\bigotimes_{j=1}^{d} \mathcal{A}_j := \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d := \sigma(p_1, \ldots, p_d)$$

heißt das *Produkt* der σ-Algebren $\mathcal{A}_1, \ldots, \mathcal{A}_d$ oder auch die *Produkt-σ-Algebra* von $\mathcal{A}_1, \ldots, \mathcal{A}_d$.

8.2 Satz *Für jedes $1 \le j \le d$ sei $\mathcal{E}_j$ ein Erzeuger der σ-Algebra $\mathcal{A}_j$ über Ω_j und $\mathcal{E}_j$ enthalte eine Folge $(E_{jk})_k$ mit $E_{jk} \uparrow \Omega_j$. Dann wird $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d$ von dem System $\mathcal{E}_1 \times \cdots \times \mathcal{E}_d$ aller Mengen $E_1 \times \cdots \times E_d$ mit $E_j \in \mathcal{E}_j$ erzeugt:*

$$\bigotimes_{j=1}^{d} \mathcal{A}_j = \sigma(\mathcal{E}_1 \times \cdots \times \mathcal{E}_d)\ .$$

Insbesondere gilt $\bigotimes_{j=1}^{d} \mathcal{A}_j = \sigma(\mathcal{A}_1 \times \cdots \times \mathcal{A}_d)$.

Beweis: Es sei $\mathcal{A}$ eine σ-Algebra über Ω. Zu zeigen ist: Jede Projektion p_j ist genau dann $\mathcal{A}/\mathcal{A}_j$-messbar, wenn $E_1 \times \cdots \times E_d \in \mathcal{A}$ ist für alle $E_j \in \mathcal{E}_j$. Nach

Satz 3.3 ist p_j genau dann $\mathcal{A}/\mathcal{A}_j$-messbar, wenn aus $E_j \in \mathcal{E}_j$ stets $p_j^{-1}(E_j) \in \mathcal{A}$ folgt. Dann liegt auch $E_1 \times \cdots \times E_d = p_1^{-1}(E_1) \cap \cdots \cap p_d^{-1}(E_d)$ in $\mathcal{A}$.

Gilt umgekehrt $E_1 \times \cdots \times E_d \in \mathcal{A}$ für alle $E_j \in \mathcal{E}_j$, so liegen für jedes gegebene $j = 1, \ldots, d$ und $E_j \in \mathcal{E}_j$ die Mengen

$$F_k := E_{1k} \times \cdots \times E_{j-1,k} \times E_j \times E_{j+1,k} \times \cdots \times E_{dk}\,, \quad k \geq 1,$$

in $\mathcal{A}$, und die Folge $(F_k)_k$ steigt gegen $\Omega_1 \times \cdots \times \Omega_{j-1} \times E_j \times \Omega_{j+1} \times \cdots \times \Omega_d = p_j^{-1}(E_j)$ auf, also ist $p_j^{-1}(E_j) \in \mathcal{A}$. □

8.3 Beispiel Wir wählen $\Omega_i := \mathbb{R}$, $\mathcal{A}_i := \mathcal{B}$ und $\mathcal{E}_i := \mathcal{R}^1$ (die Menge aller nach rechts halboffenen Intervalle in $\mathbb{R}$) für jedes $i = 1, \ldots, d$. Das System aller Mengen $E_1 \times \cdots \times E_d$ mit $E_i \in \mathcal{R}^1$ fällt zusammen mit $\mathcal{R}^d$. Da $\mathcal{B}^d = \sigma(\mathcal{R}^d)$, folgt mit Satz 8.2:

$$\mathcal{B}^d = \mathcal{B}^1 \otimes \cdots \otimes \mathcal{B}^1 \qquad (d \text{ Faktoren}).$$

8.4 Satz *Bezeichnen $(\Omega_j, \mathcal{A}_j)$, $0 \leq j \leq d$, messbare Räume, so gilt: Eine Abbildung $f : \Omega_0 \to \Omega_1 \times \cdots \times \Omega_d$ ist genau dann $\mathcal{A}_0/\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d$-messbar, wenn jede der Abbildungen $p_j \circ f$, $1 \leq j \leq d$, $\mathcal{A}_0/\mathcal{A}_j$-messbar ist.*

Beweis: Da die Komposition messbarer Abbildungen messbar ist (Satz 3.5), folgt die eine Richtung sofort. Für die Rückrichtung nutzen wir, dass

$$\mathcal{E} = \bigcup_{j=1}^{d} p_j^{-1}(\mathcal{A}_j)$$

ein Erzeuger von $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d$ ist. Jede Menge $E \in \mathcal{E}$ ist von der Form $E = p_j^{-1}(A_j)$ mit $A_j \in \mathcal{A}_j$ und $j \in \{1, \ldots, d\}$. Somit gilt $f^{-1}(E) = (p_j \circ f)^{-1}(A_j) \in \mathcal{A}_0$. Damit ist nach Satz 3.3 f $\mathcal{A}_0/\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_d$-messbar. □

Unter welchen Voraussetzungen läßt sich die Existenz eines Maßes μ auf $(\Omega, \bigotimes_{j=1}^{d} \mathcal{A}_j)$ zeigen, das der Bedingung

$$\mu(A_1 \times \cdots \times A_d) = \mu_1(A_1) \ldots \mu_d(A_d)$$

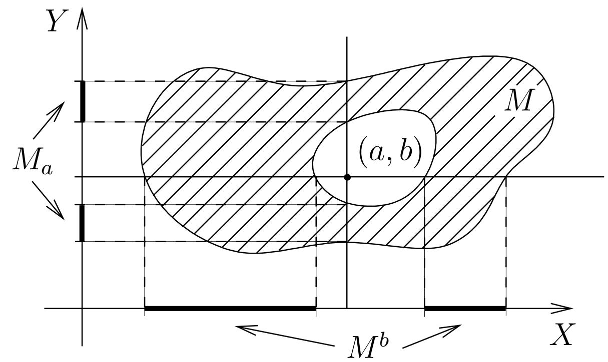
für alle $A_j \in \mathcal{A}_j$ genügt? Wir betrachten $d = 2$. Es seien fortan $(X, \mathcal{A}, \mu)$ und $(Y, \mathcal{B}, \nu)$ zwei Maßräume.

Wir führen die folgende Notation ein: für ein $M \in \mathcal{A} \otimes \mathcal{B}$ bezeichne

$$M_a := \{y \in Y : (a, y) \in M\},\ a \in X$$

$$M^b := \{x \in X : (x, b) \in M\},\ b \in Y$$

den *a-Schnitt* M_a bzw. *b-Schnitt* M^b von M oder kurz, einen *Schnitt* von M.



8.5 Lemma *Für $M \in \mathcal{A} \otimes \mathcal{B}$ ist jeder Schnitt M_a, $a \in X$, und M^b, $b \in Y$, messbar, d.h. $M_a \in \mathcal{B}$ und $M^b \in \mathcal{A}$.*

Beweis: Wir betrachten Teilmengen Q, Q_n, $n \in \mathbb{N}$, von $X \times Y$. Für $a \in X$ ist

$$\Big((X \times Y) \setminus Q\Big)_a = Y \setminus Q_a$$

und $(\bigcup_{n \geq 1} Q_n)_a = \bigcup_{n \geq 1} (Q_n)_a$. Ferner ist $(X \times Y)_a = Y$ und $(A \times B)_a = B$ bzw. $\emptyset$, je nachdem, ob a in A liegt oder nicht ($A \subset X$, $B \subset Y$). Also ist das System aller $Q \subset X \times Y$ mit $Q_a \in \mathcal{B}$ eine σ-Algebra in $X \times Y$, die alle Mengen $A \times B$ ($A \in \mathcal{A}$, $B \in \mathcal{B}$) enthält, also enthält sie auch $\mathcal{A} \otimes \mathcal{B}$. Für b-Schnitte geht der Beweis analog. □

Ein Konstruktionsweg des gewünschten Maßes geht auf B. Cavalieri (1591-1647), ein Schüler von G. Galilei, zurück. Das Cavalieri*sche Prinzip* ist, das Maß (Volumen) einer Menge „durch Zerlegen dieser Menge in dünne parallele Scheiben und kontinuierliches Aufsummieren“ (Integrieren) der Volumina dieser Scheiben zu bestimmen. Dieses Prinzip wollen wir nun mathematisch präzisieren.

8.6 Lemma *Ist das Maß ν auf $(Y, \mathcal{B})$ σ-endlich, so ist für jedes $M \in \mathcal{A} \otimes \mathcal{B}$ die Funktion $x \mapsto \nu(M_x)$ auf X definiert und $\mathcal{A}$-messbar.*

Beweis: Es sei

$$f_M \colon X \to \overline{\mathbb{R}},\ f_M(x) := \nu(M_x),\ x \in X,\ M \in \mathcal{A} \otimes \mathcal{B}\ .$$

Sei zunächst $\nu(Y) < \infty$. Setze

$$\mathcal{M} := \{M \in \mathcal{A} \otimes \mathcal{B} \colon f_M \text{ ist } \mathcal{A}\text{-messbar}\}.$$

Für $A \in \mathcal{A}$ und $B \in \mathcal{B}$ ist $f_{A \times B}(x) = \nu(B) 1_A(x)$, $x \in X$, also $\mathcal{A}$-messbar, also gilt $\{A \times B,\ A \in \mathcal{A},\ B \in \mathcal{B}\} \subset \mathcal{M}$. Da das System aller Mengen $A \times B$

durchschnittstabil ist und nach Satz 8.2 $\mathcal{A}\otimes\mathcal{B}$ erzeugt, folgt $\mathcal{M}=\mathcal{A}\otimes\mathcal{B}$, wenn wir zeigen können, dass $\mathcal{M}$ ein DYNKIN-System ist (DYNKIN-System Argument). Nun ist $X\times Y\in\mathcal{M}$, denn $f_{X\times Y}$ ist konstant gleich $\nu(Y)$, also $\mathcal{A}$-messbar. Mit $M\in\mathcal{M}$ ist auch $M^c\in\mathcal{M}$, denn für $x\in X$ ist

$$f_{M^c}(x)=\nu((M_x)^c)=\nu(Y)-f_M(x),$$

da $\nu(Y)<\infty$. Also ist f_{M^c} $\mathcal{A}$-messbar. Ist nun $(M_n)_n$ eine disjunkte Folge in $\mathcal{M}$, so ist auch $\bigcup_{n\geq 1}M_n\in\mathcal{M}$, denn $f_{\bigcup_{n\geq1}M_n}=\sum_{n\geq1}f_{M_n}$ ist $\mathcal{A}$-messbar. Ist ν nur σ-endlich, so existiert eine Folge $(B_n)_n$ von Mengen aus $\mathcal{B}$ mit $B_n\uparrow Y$ und $\nu(B_n)<\infty$, $n\in\mathbb{N}$. Es sei $\nu_n\colon\mathcal{B}\to\overline{\mathbb{R}}$ gegeben durch $\nu_n(B):=\nu(B\cap B_n)$, $B\in\mathcal{B}$, $n\in\mathbb{N}$. Für $M\in\mathcal{A}\otimes\mathcal{B}$ ist $x\mapsto\nu_n(M_x)$ $\mathcal{A}$-messbar, also ist auch $f_M(x)=\lim_{n\to\infty}\nu_n(M_x)$ eine $\mathcal{A}$-messbare Funktion. □

Jetzt ergibt sich die Existenz des gesuchten Maßes wie folgt:

8.7 Satz *Sind μ und ν σ-endlich, so gibt es genau ein Maß $\mu\otimes\nu\colon\mathcal{A}\otimes\mathcal{B}\to[0,\infty]$ mit $\mu\otimes\nu(A\times B)=\mu(A)\,\nu(B)$, $A\in\mathcal{A}$, $B\in\mathcal{B}$, und es ist*

$$\mu\otimes\nu(M)=\int_X\nu(M_x)\,d\mu(x)=\int_Y\mu(M^y)\,d\nu(y)$$

für $M\in\mathcal{A}\otimes\mathcal{B}$. Das Maß $\mu\otimes\nu$ ist σ-endlich.

Beweis: Für jedes $M\in\mathcal{A}\otimes\mathcal{B}$ ist $f_M\geq 0$ und mit Lemma 8.6 ist

$$\varrho(M):=\int_X\nu(M_x)\,d\mu(x)$$

definiert und es gilt offenbar $\varrho(A\times B)=\nu(B)\,\mu(A)$. Wir müssen nun die σ-Additivität zeigen. Es sei $(M_n)_n$ eine Folge disjunkter Mengen in $\mathcal{A}\otimes\mathcal{B}$, so ist

$$\begin{aligned}\varrho\Big(\bigcup_{n\geq1}M_n\Big) &= \int_X\nu\Big(\big(\bigcup_{n\geq1}M_n\big)_x\Big)\,d\mu(x)=\int_X\sum_{n\geq1}\nu\big((M_n)_x\big)\,d\mu(x)\\ &= \sum_{n\geq1}\int_X\nu\big((M_n)_x\big)\,d\mu(x)=\sum_{n\geq1}\varrho(M_n)\ .\end{aligned}$$

Analog ist nun auch $\sigma:\mathcal{A}\otimes\mathcal{B}\to[0,\infty]$ mit

$$\sigma(M):=\int_Y\mu(M^y)\,d\nu(y),\quad M\in\mathcal{A}\otimes\mathcal{B}$$

ein Maß auf $\mathcal{A}\otimes\mathcal{B}$ mit $\sigma(A\times B)=\mu(A)\,\nu(B)$, $A\in\mathcal{A}$, $B\in\mathcal{B}$. Es kann aber höchstens ein derartiges Maß geben, denn das System aller Mengen $A\times B$ ist durchschnittstabiler Erzeuger von $\mathcal{A}\otimes\mathcal{B}$ und erfüllt somit alle Voraussetzungen des Eindeutigkeitssatzes, Satz 2.9. Weiterhin ergibt sich aus der Eigenschaft $\mu \otimes \nu(A \times B) = \mu(A)\,\nu(B)$ auch die σ-Endlichkeit von $\mu \otimes \nu$. Damit ist der Satz bewiesen. $\square$

8.8 Definition Das Maß $\mu \otimes \nu$ heißt das *Produktmaß* von μ und ν.

8.9 Beispiel Das Beispiel 8.3 lehrt $\lambda^2 = \lambda^1 \otimes \lambda^1$, und allgemeiner $\lambda^p \otimes \lambda^q = \lambda^{p+q}$.

8.10 Corollary (Cavalierisches Prinzip) *Es seien μ, ν σ-endlich und für $M, N \in \mathcal{A} \otimes \mathcal{B}$ sei $\nu(M_x) = \nu(N_x)$ für μ-fast alle $x \in X$. Dann ist $\mu \otimes \nu(M) = \mu \otimes \nu(N)$.*

Wir bereiten nun den Satz von Fubini vor:

8.11 Lemma *Sei $(Z, \mathcal{C})$ ein weiterer Messraum. Für jede messbare Abbildung $f\colon (X \times Y, \mathcal{A} \otimes \mathcal{B}) \to (Z, \mathcal{C})$ sind alle Schnitte $f(a, \cdot)\colon (Y, \mathcal{B}) \to (Z, \mathcal{C})$, $a \in X$, und $f(\cdot, b)\colon (X, \mathcal{A}) \to (Z, \mathcal{C})$, $b \in Y$, messbar.*

Beweis: Sei $C \in \mathcal{C}$, $a \in X$, $b \in Y$, so gilt

$$\big(f(a,\cdot)\big)^{-1}(C) = \big(f^{-1}(C)\big)_a \text{ und } \big(f(\cdot,b)\big)^{-1}(C) = \big(f^{-1}(C)\big)^b ,$$

also ist die Behauptung mit Hilfe von Lemma 8.5 bewiesen. $\square$

8.12 Satz (von Fubini) *Es seien μ, ν σ-endlich. Dann gilt*

(a) Für jede nicht-negative $\mathcal{A} \otimes \mathcal{B}$-messbare numerische Funktion f sind die durch

$$x \mapsto \int_Y f(x,y)\, d\nu(y) \quad \text{bzw.} \quad y \mapsto \int_X f(x,y)\, d\mu(x)$$

auf X bzw. Y definierten nicht-negativen numerischen Funktionen $\mathcal{A}$-messbar bzw. $\mathcal{B}$-messbar und es gilt

$$\int_{X\times Y} f\, d\mu \otimes \nu = \int_X \Big(\int_Y f(x,y)\, d\nu(y) \Big)\, d\mu(x) = \int_Y \Big(\int_X f(x,y)\, d\mu(x) \Big)\, d\nu(y) \,. \tag{8.1}$$

(b) Ist $f\colon X \times Y \to \overline{\mathbb{R}}$ $\mu \otimes \nu$-integrierbar, so ist $f(x, \cdot)$ ν-integrierbar für μ-fast alle $x \in X$ und $f(\cdot, y)$ ist μ-integrierbar für ν-fast alle $y \in Y$. Die Funktionen $x \mapsto \int_Y f(x,y)\, d\nu(y)$ bzw. $y \mapsto \int_X f(x,y)\, d\mu(x)$ sind somit μ- bzw. ν- fast überall definiert und μ- bzw. ν-integrierbar und es gilt (8.1).

Beweis: zu (a): Für $M \in \mathcal{A} \otimes \mathcal{B}$ ist $M_x \in \mathcal{B},\ x \in X$. Die Funktion

$$x \mapsto \nu(M_x) = \int_Y 1_M(x,y)\,d\nu(y)$$

ist $\mathcal{A}$-messbar und nach Satz 8.7 ist

$$\int_{X\times Y} 1_M\,d\mu\otimes\nu = \int_X \Big(\int_Y 1_M(x,y)\,d\nu(y)\Big)\,d\mu(x)\ .$$

Analog mit vertauschten Rollen für μ und ν. Also gilt (a) für $f = 1_M$ mit $M \in \mathcal{A}\otimes\mathcal{B}$ und damit für jedes $f \in \mathcal{EF}(X\times Y, \mathbb{R}^+)$. Für eine nicht-negative $\mathcal{A}\otimes\mathcal{B}$-messbare numerische Funktion f gibt es eine Folge $(f_n)_n$ in $\mathcal{EF}(X\times Y,\mathbb{R}^+)$ mit $f_n \uparrow f$. Für $x \in X$ ist $f(x,\cdot) \in \mathcal{L}_0(Y,\mathcal{B},\overline{\mathbb{R}}^+)$, siehe Lemma 8.11, und $f_n(x,\cdot) \in \mathcal{EF}(Y,\mathbb{R}^+)$ und $f_n(x,\cdot) \uparrow f(x,\cdot)$. Nach der Definition des Integrals folgt für $x \in X$

$$\int_Y f_n(x,y)\,d\nu(y) \uparrow \int_Y f(x,y)\,d\nu(y), \tag{8.2}$$

und auf der linken Seite steht eine Folge $\mathcal{A}$-messbarer Funktionen in x und rechts steht somit ebenfalls eine $\mathcal{A}$-messbare Funktion in x. Somit ist

$$\begin{aligned}\int_{X\times Y} f\,d\mu\otimes\nu &= \lim_{n\to\infty}\int_{X\times Y} f_n\,d\mu\otimes\nu = \lim_{n\to\infty}\int_X\Big(\int_Y f_n(x,y)\,d\nu(y)\Big)\,d\mu(x)\\ &= \int_X\Big(\lim_{n\to\infty}\int_Y f_n(x,y)\,d\nu(y)\Big)\,d\mu(x) = \int_X\Big(\int_Y f(x,y)\,d\nu(y)\Big)\,d\mu(x)\ .\end{aligned}$$

Dabei benutzen wir die Definition des Integrals, (8.1) für einfache Funktionen sowie monotone Konvergenz und (8.2). Eine analoge Herleitung mit vertauschten Rollen für μ und ν liefert den Beweis zu (a).

zu (b): Mit f ist auch $|f|$ integrierbar bzgl. $\mu\otimes\nu$ und (a) besagt

$$\int_X\Big(\int_Y |f(x,y)|\,d\nu(y)\Big)\,d\mu(x) = \int_{X\times Y}|f|\,d\mu\otimes\nu < \infty\ .$$

Die Funktion in der Klammer der linken Seite ist nach (a) eine $\mathcal{A}$-messbare numerische Funktion von $x \in X$ und die Endlichkeit des Integrals impliziert

$$\int_Y |f(x,y)|\,d\nu(y) < \infty$$

für μ-fast alle $x \in X$. Für μ-fast alle x gilt somit

$$\int_Y f(x,y)\,d\nu(y) = \int_Y f^+(x,y)\,d\nu(y) - \int_Y f^-(x,y)\,d\nu(y)\ . \tag{8.3}$$

Nach (a) sind die Integrale der rechten Seite als Funktionen in $x \in X$ nicht-negative $\mathcal{A}$-messbare numerische Funktionen, und sie sind μ-integrierbar, denn

$$\int_X\Big(\int_Y f^{\pm}(x,y)\,d\nu(y)\Big)\,d\mu(x) \le \int_X\Big(\int_Y |f(x,y)|\,d\nu(y)\Big)\,d\mu(x) < \infty\ .$$

Damit ist 8.3 als Funktion in x μ-fast überall integrierbar und

$$\begin{aligned}
&\int_X \Big(\int_Y f(x,y)\,d\nu(y)\Big)\,d\mu(x)\\
&\qquad = \int_X \Big(\int_Y f^+(x,y)\,d\nu(y)\Big)\,d\mu(x) - \int_X \Big(\int_Y f^-(x,y)\,d\nu(y)\Big)\,d\mu(x)\\
&\qquad = \int_{X\times Y} f^+\,d\mu\otimes\nu - \int_{X\times Y} f^-\,d\mu\otimes\nu = \int_{X\times Y} f\,d\mu\otimes\nu\ .
\end{aligned}$$

Analog schließt man bei vertauschten Rollen für μ und ν. Somit ist der Satz bewiesen. □

8.13 Beispiele

(a) Es seien $\lambda^2 = \lambda^1 \otimes \lambda^1$, $A := \mathbb{Q}\times\mathbb{R} \in \mathcal{B}^2$ und $f := 1_A$, dann ist $\lambda^2(A) = \int f d\lambda^2 = 0$ nach Satz 8.7, also ist f λ^2-integrierbar. Aber $f(x,\cdot)$ ist für $x \in \mathbb{Q}$ nicht λ^1-integrierbar.

(b) Gegeben sei die Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R} \quad \text{mit} \quad f(x,y) := \begin{cases} \frac{xy}{(x^2+y^2)^2} & ,\ (x,y) \neq (0,0) \\ 0 & ,\ (x,y) = (0,0) \end{cases} \ .$$

Dann ist f eine $\mathcal{B}^2$-messbar Funktion und für jedes $y \in \mathbb{R}$ konvergiert das uneigentliche Cauchy-Riemann Integral $\int_{\mathbb{R}} f(x,y)\,dx$ absolut. Weiter ist $f(\cdot,y)$ ungerade, also

$$\int_{\mathbb{R}} f(x,y)\,d\lambda^1(x) = 0,$$

und wegen $f(x,y) = f(y,x)$ ist auch

$$\int_{\mathbb{R}} \Big(\int_{\mathbb{R}} f(x,y)\,d\lambda^1(x)\Big)\,d\lambda^1(y) = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} f(x,y)\,d\lambda^1(y)\Big)\,d\lambda^1(x) = 0\ .$$

Wäre f λ^2-integrierbar, so wäre nach dem Satz von Fubini auch $x \mapsto \int_{\mathbb{R}} |f(x,y)|\,d\lambda^1(y)$ λ^1-integrierbar, aber $\int_{\mathbb{R}} \frac{|xy|}{(x^2+y^2)^2}\,d\lambda^1(y) = \frac{1}{|x|}$ für $x \neq 0$, also ist f nicht integrierbar, d.h. $f \in \mathcal{L}^0(\mathbb{R}^2) \setminus \mathcal{L}^1(\mathbb{R}^2)$. Aus der Existenz und Gleichheit der iterierten Integrale kann nicht auf die Integrierbarkeit von f geschlossen werden.

(c) Für $x, y > 0$ ist $\frac{x^2-y^2}{(x^2+y^2)^2} = \frac{\partial^2}{\partial x \partial y} \arctan\frac{x}{y}$, also ist

$$\int_0^1 \Big(\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,d\lambda^1(y)\Big)\,d\lambda^1(x) = \frac{\pi}{4}$$

und

$$\int_0^1 \Big(\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,d\lambda^1(x)\Big)\,d\lambda^1(y) = -\frac{\pi}{4}.$$

Die iterierten Integrale existieren beide, sind aber nicht gleich, also ist die Funktion nicht λ^2-integrierbar über $(0,1)^2$ (Beispiel von Cauchy 1814).

(d) Es seien $(X,\mathcal{A},\mu) = (Y,\mathcal{B},\nu) = (\mathbb{N},\mathcal{P}(\mathbb{N}),\varrho)$, wobei ϱ das Zählmaß auf $\mathbb{N}$ ist. Es ist $\mathcal{P}(\mathbb{N})\otimes\mathcal{P}(\mathbb{N}) = \mathcal{P}(\mathbb{N}\times\mathbb{N})$ und der Satz von Fubini sagt

$$\sum_{m,n\in\mathbb{N}} a_{mn} = \sum_{m\geq 1}\sum_{n\geq 1} a_{mn} = \sum_{n\geq 1}\sum_{m\geq 1} a_{mn}$$

für alle $a_{mn}\in[0,\infty]$ und die Gleichheit gilt auch für $a_{mn}\in\mathbb{R}$, falls eine der Reihen bei Ersetzung von a_{mn} durch $|a_{mn}|$ konvergiert. Dies ist der (große) Umordnungssatz für Doppelreihen.

(e) Es ist

$$\int_0^\infty \Big(\int_0^\infty y e^{-(1+x^2)y^2}\,dy\Big)dx = \frac{1}{2}\int_0^\infty \frac{dx}{1+x^2}\ ,$$

denn $-\frac{1}{2(1+x^2)}e^{-(1+x^2)y^2}\Big|_0^\infty = \frac{1}{2(1+x^2)}$. Nun ist $\frac{1}{2}\int_0^\infty \frac{dx}{1+x^2} = \frac{1}{2}\arctan x\Big|_0^\infty = \frac{\pi}{4}$. Andererseits ist

$$\int_0^\infty\Big(\int_0^\infty e^{-x^2y^2}\,dx\Big)ye^{-y^2}\,dy = \int_0^\infty\Big(\int_0^\infty e^{-t^2}\,dt\Big)e^{-y^2}\,dy = \Big(\int_0^\infty e^{-y^2}\,dy\Big)^2.$$

Da der Integrand nicht-negativ ist, können wir die Integrationsreihenfolge vertauschen und es folgt $\int_0^\infty e^{-x^2}dx = \frac{1}{2}\sqrt{\pi}$. Eine einfache Substitution liefert

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{+\infty} e^{-(x-m)^2/2\,\sigma^2}\,dx = 1, \quad m\in\mathbb{R}, \sigma>0.$$

Dieser Beweis stammt von Laplace (1778).

Wir können $(\Omega_1\times\cdots\times\Omega_{d-1})\times\Omega_d$ und $\Omega_1\times\cdots\times\Omega_d$ mittels $((\omega_1,\dots\omega_{d-1}),\,\omega_d)\mapsto(\omega_1,\dots,\omega_d)$ identifizieren und erhalten damit auch

$$(\mathcal{A}_1\otimes\cdots\otimes\mathcal{A}_{d-1})\otimes\mathcal{A}_d = \mathcal{A}_1\otimes\cdots\otimes\mathcal{A}_d\,. \tag{8.4}$$

Es gilt sogar

$$\Big(\bigotimes_{j=1}^k \mathcal{A}_j\Big)\otimes\Big(\bigotimes_{j=k+1}^d \mathcal{A}_j\Big) = \bigotimes_{j=1}^d \mathcal{A}_j \quad \text{für alle } 1\leq k\leq d\,.$$

Mittels (8.4) kann dann die Existenz des Produktmaßes für beliebige $d\geq 2$ per Induktion bewiesen werden.

Es gilt also:

8.14 Satz *Zu gegebenen σ-endlichen Maßen $\mu_1, \dots, \mu_d$ auf $(\Omega_1, \mathcal{A}_1), \dots, (\Omega_d, \mathcal{A}_d)$ gibt es genau ein Maß μ auf $(\prod_{j=1}^d \Omega_j, \bigotimes_{j=1}^d \mathcal{A}_j)$ mit*

$$\mu(A_1 \times \cdots \times A_d) = \mu_1(A_1) \dots \mu_d(A_d)$$

für alle $A_j \in \mathcal{A}_j$, $j = 1, \dots, d$. μ ist σ-endlich und heißt Produkt der Maße $\mu_1, \dots, \mu_d$ bzw. Produktmaß und wird mit

$$\bigotimes_{j=1}^d \mu_j = \mu_1 \otimes \cdots \otimes \mu_d$$

bezeichnet. Es gilt

$$\Big(\bigotimes_{j=1}^k \mu_j\Big) \otimes \Big(\bigotimes_{j=k+1}^d \mu_j\Big) = \bigotimes_{j=1}^d \mu_j\,, \quad 1 \leq k \leq d\,. \tag{8.5}$$

(Assoziativität)

Mit (8.5) und einer Induktion über d kann man den Satz von Fubini übertragen (Übung!).

Wir betrachten nun den Fall, in dem jedes der Maße μ_j mit einer reellen Dichte $f_j \geq 0$ versehen wird. Dann ist mit μ_j auch $\nu_j := f_j \mu_j$ σ-endlich (dies beweisen wir hier nicht, da unsere Maße immer W-Maße sind) und es gilt:

8.15 Satz *Für jedes $j = 1, \dots, d$ seien $(\Omega_j, \mathcal{A}_j, \mu_j)$ σ-endliche Maßräume, $f_j \geq 0$ reelle $\mathcal{A}_j$-messbare numerische Funktionen auf Ω_j und $\nu_j := f_j \mu_j$. Dann ist das Produkt $\nu_1 \otimes \cdots \otimes \nu_d$ definiert und es gilt*

$$\nu_1 \otimes \cdots \otimes \nu_d = (f_1 \otimes \cdots \otimes f_d) \cdot (\mu_1 \otimes \cdots \otimes \mu_d)$$

mit

$$f_1 \otimes \cdots \otimes f_d(\omega_1, \dots, \omega_d) := f_1(\omega_1) \cdots f_d(\omega_d)$$

(Tensorprodukt von $f_1, \dots, f_d$).

Beweis: Wir beweisen hier nur den Fall $d = 2$, der Rest folgt per Induktion. Es seien $A_i \in \mathcal{A}_i$, $i = 1, 2$, dann ist

$$\begin{aligned}
\nu_1(A_1)\nu_2(A_2) &= \Big(\int_{A_1} f_1 \, d\mu_1\Big)\Big(\int_{A_2} f_2 \, d\mu_2\Big) \\
&= \iint 1_{A_1}(\omega_1) f_1(\omega_1)\, 1_{A_2}(\omega_2) f_2(\omega_2)\, \mu_1(d\omega_1)\mu_2(d\omega_2) \\
&= \iint 1_{A_1 \times A_2}(\omega_1, \omega_2)\, f_1 \otimes f_2(\omega_1, \omega_2)\, \mu_1(d\omega_1)\, \mu_2(d\omega_2) \\
&\overset{\text{Fubini}}{=} \int_{A_1 \times A_2} f_1 \otimes f_2 \, d(\mu_1 \otimes \mu_2).
\end{aligned}$$

Es folgt mit Satz 8.7
$$\nu_1 \otimes \nu_2 = (f_1 \otimes f_2)\,(\mu_1 \otimes \mu_2)\,. \qquad \square$$

Die folgende Anwendung des Satzes von FUBINI ist sehr nützlich. Sie erlaubt in bestimmten Situationen μ-Integrale durch LEBESGUE-Integrale auszudrücken:

8.16 Satz *Es sei* $(\Omega,\, \mathcal{A},\, \mu)$ *ein* σ*-endlicher Maßraum und* f *eine nicht-negative* $\mathcal{A}$*-messbare reelle Funktion. Ferner sei* $\varphi : [0,\, \infty) \to [0,\, \infty)$ *eine monoton wachsende stetige Funktion mit* $\varphi(0) = 0$*, die auf* $(0,\, \infty)$ *stetig differenzierbar ist. Dann gilt für alle* $A \in \mathcal{A}$
$$\begin{aligned}\int_A \varphi \circ f \, d\mu &= \int_{(0,\,\infty)} \varphi'(t)\, \mu\big(\{f \geq t\} \cap A\big)\, \lambda(dt)\\ &= \int_0^\infty \varphi'(t)\, \mu\big(\{f \geq t\} \cap A\big) dt\,.\end{aligned}$$

Beweis: Sei μ zunächst endlich. Dann kann ohne Einschränkung $A = \Omega$ gesetzt werden, denn mit
$$\tilde{\mu} := \mu(\,\cdot\, \cap A)$$
gilt
$$\int_A \varphi \circ f \, d\mu = \int \varphi \circ f \, d\tilde{\mu}$$
und
$$\mu\big(\{f \geq t\} \cap A\big) = \tilde{\mu}(f \geq t)$$
für alle $A \in \mathcal{A}$. Sei λ^* das auf $(0,\, \infty) \cap \mathcal{B}^1$ definierte Lebesgue-Maß. Da φ monoton wachsend ist, gilt $\varphi'(t) \geq 0$ für alle $t > 0$. Die stetige Funktion φ' ist wegen $[\frac{1}{n},\, a] \uparrow (0,\, a]$ über jedem Intervall $(0,\, a]$, $a > 0$, λ^*-integrierbar und
$$\int_{(0,\,a]} \varphi'(t)\, \lambda^*(dt) = \lim_{n\to\infty} \int_{\frac{1}{n}}^a \varphi'(t)\, dt = \varphi(a) - \lim_{n\to\infty} \varphi(1/n) = \varphi(a)\,,$$
da $\varphi(0) = 0$ und φ in $t = 0$ stetig. Aus $f \geq 0$ folgt
$$\int_{(0,\, f(\omega)]} \varphi'(t)\, \lambda^*(dt) = \varphi(f(\omega))\,, \quad \omega \in \Omega\,.$$
Es folgt
$$\begin{aligned}\int \varphi \circ f \, d\mu &= \int \Big(\int_{(0,\, f(\omega)]} \varphi'(t)\, \lambda^*(dt)\Big)\, \mu(d\omega)\\ &= \iint \varphi'(t)\, 1_{(0,\, f(\omega)]}(t)\, \lambda^*(dt)\, \mu(d\omega)\\ &= \iint \varphi'(t)\, 1_E(\omega,\, t)\, \lambda^*(dt)\, \mu(d\omega)\end{aligned}$$

mit $E := \{(\omega, t) \in \Omega \times (0, \infty) \,:\, f(\omega) \geq t\}$. Die letzte Gleichheit folgt, da $E \in \mathcal{A} \otimes \big(\mathcal{B}^1 \cap (0, +\infty)\big)$. Wenn dies geklärt ist, liefert der Satz von FUBINI

$$\begin{aligned}\int \varphi \circ f \, d\mu &= \iint \varphi'(t)\, 1_E(\omega, t)\, \mu(d\omega)\, \lambda^*(dt) \\ &= \int \varphi'(t)\, \mu(E_t)\, \lambda^*(dt) = \int \varphi'(t)\, \mu\big(\{f \geq t\}\big)\, \lambda^*(dt)\,,\end{aligned}$$

denn der t-Schnitt von E ist die Menge aller ω mit $f(\omega) \geq t$. Da $t \mapsto \mu\big(\{f \geq t\}\big)$ linksseitig stetig und monoton fallend ist, hat die Funktion höchstens abzählbar unendlich viele Unstetigkeitsstellen auf $(0, \infty)$. Zusammen mit der Stetigkeit von φ' liefert dies die uneigentliche RIEMANN-Integrierbarkeit von $t \mapsto \varphi'(t)\, \mu\big(\{f \geq t\}\big)$ auf $(0, \infty)$, wobei der Wert ∞ zugelassen ist.

Zu zeigen bleibt: $E \in \mathcal{A} \otimes \big(\mathcal{B}^1 \cap (0, \infty)\big)$. Dies ist der Inhalt einer Übungsaufgabe.

Ist μ σ-endlich, so folgt die Aussage einfach mit dem Satz von der monotonen Konvergenz. Dies ist eine Übung. $\square$

8.17 Bemerkung Die Aussage gilt analog, wenn man $\mu\big(\{f \geq t\}\big)$ durch $\mu\big(\{f > t\}\big)$ ersetzt, denn $t \mapsto \mu\big(\{f > t\}\big)$ hat höchstens abzählbar viele Unstetigkeitsstellen, also ist

$$\mu\big(\{f > t\}\big) = \mu\big(\{f \geq t\}\big) \quad \lambda\text{-f.ü.}$$

8.18 Beispiel Es sei $\varphi(t) = t^p$, $p > 0$. Für jede $\mathcal{A}$-messbare reelle Funktion $f \geq 0$ folgt

$$\int f^p \, d\mu = p \int_0^\infty t^{p-1}\, \mu\big(\{f \geq t\}\big)\, dt$$

und für $p = 1$

$$\int f \, d\mu = \int_{\mathbb{R}^+} \mu\big(\{f \geq t\}\big)\, \lambda^1(dt) = \int_0^\infty \mu\big(\{f \geq t\}\big)\, dt\,.$$

Das Integral $\int f \, d\mu$ wird “vertikal”, das Integral auf der rechten Seite jedoch “horizontal” gebildet.

Ist X eine nicht-negative Zufallsgröße auf einem W-Raum $(\Omega, \mathcal{A}, P)$, so folgt insbesondere

$$\begin{aligned}\mathbb{E}(X) &= \int_0^\infty P(X > t)\, dt \\ \mathbb{E}(X^p) &= \int_0^\infty p\, t^{p-1}\, P(X > t)\, dt\,, \quad p > 0\,.\end{aligned}$$

Aus der letzten Gleichung folgt

$$\sum_{n\geq 1} p(n-1)^{p-1}\,P(X>n) \leq \mathbb{E}(X^p)$$
$$\leq \sum_{n\geq 0} p(n+1)^{p-1}\,P(X>n)\,, \quad p\geq 1\,.$$

Ist X eine Zufallsgröße und existiert $\mathbb{E}(X)$, so folgt durch Zerlegung von X in Positiv- und Negativteil

$$\mathbb{E}(X) = \int_0^\infty \big(P(X>t) - P(X<-t)\big)\,dt\,.$$

Existiert auch $\mathrm{Var}(X)$, so gilt

$$\mathrm{Var}(X) = \int_0^\infty \big((2t-\mathbb{E}(X))P(X>t) + (-2t+\mathbb{E}(X))P(X<-t)\big)\,dt.$$

Satz 8.15 liefert uns zu d Experimenten $(\Omega_j,\,\mathcal{A}_j,\,\mu_j)$, $j=1,\dots,d$, einen gemeinsamen W-Raum $(\prod\Omega_j,\,\bigotimes\mathcal{A}_j,\,\bigotimes\mu_j)$, der die Ausführung der d Einzelexperimente beschreibt. Die Wahl des Produktmaßes bekommt die Interpretation stochastisch unabhängig ausgeführter Teilexperimente. Zunächst wenden wir uns aber der Frage nach der Konstruktion eines W-Raumes für unendlich viele Experimente zu.

Es sei I eine nichtleere Indexmenge und $(\Omega_i,\,\mathcal{A}_i,\,P_i)_{i\in I}$ eine Familie von W-Räumen. Für $K\subset I$ setzen wir

$$\Omega_K := \prod_{i\in K}\Omega_i\,, \quad \Omega := \prod_{i\in I}\Omega_i\,.$$

Ω_K ist die Menge aller Abbildungen $\omega: K\to\bigcup_{i\in K}\Omega_i$ mit $\omega(i)\in\Omega_i$ für alle $i\in K$. Wir restringieren diese Abbildung auf eine nichtleere Teilmenge $J\subset K$ und erhalten die Projektionsabbildung

$$p_J^K : \Omega_K \to \Omega_J\,.$$

Für $K=I$ setzen wir $p_J := p_J^I$ und $p_i^K := p_{\{i\}}^K$ für $J=\{i\}$, speziell $p_i := p_i^I$.

Es gilt

$$p_J^L = p_J^K \circ p_K^L\,, \quad J\subset K\subset L\,,$$

insbesondere

$$p_J = p_J^K \circ p_K\,, \quad J\subset K\,.$$

Es sei $\mathcal{H}=\mathcal{H}(I)$ das System aller *nichtleeren, endlichen* Teilmengen von I. Für $J\in\mathcal{H}$ sind

$$\mathcal{A}_J := \bigotimes_{i\in J}\mathcal{A}_i, \quad P_J := \bigotimes_{i\in J}P_i$$

definiert. Als Produkt-σ-Algebra für unendlich viele σ-Algebren wählen wir mit Definition 8.1

$$\mathcal{A} := \bigotimes_{i \in I} \mathcal{A}_i := \sigma\big((p_i)_{i \in I}\big) .$$

Für jedes $J \in \mathcal{H}$ ist dann p_J $\mathcal{A}/\mathcal{A}_J$-messbar, denn $\mathcal{A}_J = \sigma(p_i^J,\, i \in J)$ und $p_i = p_i^J \circ p_J$ für jedes $i \in J$ (wir verwenden also Satz 8.4). Es gilt also

$$\sigma\big((p_i)_{i \in I}\big) = \sigma(p_J,\, J \in \mathcal{H}) .$$

Welche Eigenschaft erwarten wir von einem Maß P auf $(\Omega, \mathcal{A})$? Für $A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n$ (bei einer Folge von W-Räumen $(\Omega_n, \mathcal{A}_n, P_n)$) soll $A = A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \ldots$ in $\mathcal{A}$ liegen und $P(A) = P_1(A_1) \cdots P_n(A_n)$ gelten. Allgemein wünschen wir also

$$P\Big(p_J^{-1}\big(\prod_{i \in J} A_i\big)\Big) = \prod_{i \in J} P_i(A_i) , \quad J \in \mathcal{H},\, A_i \in \mathcal{A}_i,\, i \in J .$$

Also soll $p_J(P)$ gleich P_J sein für $J \in \mathcal{H}$, denn P_J ist das einzige derartige Maß (siehe Satz 8.7).

Gibt es ein W-Maß P auf $\mathcal{A}$ derart, dass dessen Bild unter jeder Projektion p_J mit $J \in \mathcal{H}$ gleich P_J ist?

8.19 Satz (von Andersen und Jessen) *Auf der σ-Algebra $\mathcal{A} := \bigotimes_{i \in I} \mathcal{A}_i$ existiert genau ein Maß P derart, dass für jede Menge $J \in \mathcal{H}(I)$ gilt*

$$p_J(P) = P_J .$$

P ist ein W-Maß.

Der Beweis ist technisch und soll im Rahmen dieser Vorlesung nicht diskutiert werden. Für das Selbststudium ist dennoch ein Beweis angegeben.

Ein paar Vorbetrachtungen:

Für $J, K \in \mathcal{H}$, $J \subset K$, ist p_J^K $\mathcal{A}_K/\mathcal{A}_J$-messbar, denn die Mengen $\prod_{i \in J} A_i$, $A_i \in \mathcal{A}_i$, $i \in J$, erzeugen $\mathcal{A}_J$ und $(p_J^K)^{-1}\big(\prod_{i \in J} A_i\big) = \prod_{i \in K} A_i'$ mit $A_i' = A_i$ für $i \in J$ und $A_i' = \Omega$ für $i \in K \backslash J$.

Es gilt

$$\prod_{i \in K} P_i(A_i') = \prod_{i \in J} P_i(A_i),$$

also mit Satz 8.14

$$p_J^K(P_K) = P_J , \quad J \subset K,\, J, K \in \mathcal{H} .$$

Wir setzen $\mathcal{Z}_J := p_J^{-1}(\mathcal{A}_J)$, $J \in \mathcal{H}$. Diese σ-Algebra heißt die σ-Algebra der *J-Zylindermengen*. Es ist $(p_J^K)^{-1}(\mathcal{A}_J) \subset \mathcal{A}_K$. Mit $p_J = p_J^K \circ p_K$ folgt

$$\mathcal{Z}_J \subset \mathcal{Z}_K\,, \quad J \subset K,\ J, K \in \mathcal{H}\,. \tag{8.6}$$

Sei

$$\mathcal{Z} := \bigcup_{J \in \mathcal{H}} \mathcal{Z}_J.$$

Wir nennen $\mathcal{Z}$ das *System der Zylindermengen*. Je zwei Zylindermengen $Z_1, Z_2 \in \mathcal{Z}$ liegen in ein und derselben σ-Algebra $\mathcal{Z}_J$ für ein $J \in \mathcal{H}$. Für $Z_i \in \mathcal{Z}_{J_i}$, $i = 1, 2$, ist $J_1 \cup J_2$ nach (8.6) geeignet. $\mathcal{Z}$ ist also eine Algebra, i.a. jedoch keine σ-Algebra. Es gilt aber

$$\mathcal{A} = \sigma(\mathcal{Z})$$

nach Definition und $\mathcal{A} = \sigma(p_J,\ J \in \mathcal{H})$.

Nun kommen wir zum Beweis von Satz 8.19 in vier Schritten.

Beweis von Satz 8.19: 1. Schritt: Um $p_J(P) = P_J$ zu erreichen, muss das gesuchte P auf $Z = p_J^{-1}(A)$ den Wert $P_J(A)$ bekommen, $J \in \mathcal{H}$, $A \in \mathcal{A}_J$. Dieser Wert darf nur von Z, nicht von der speziellen Darstellung $Z = p_J^{-1}(A)$ abhängen!

Es sei $Z = p_J^{-1}(A) = p_K^{-1}(B)$, $J, K \in \mathcal{H}$, $A \in \mathcal{A}_J$, $B \in \mathcal{A}_K$. Wenn $J \subset K$ ist, so gilt

$$p_J^{-1}(A) = p_K^{-1}\big((p_J^K)^{-1}(A)\big)\,,$$

also

$$p_K^{-1}(B) = p_K^{-1}(B')$$

mit $B' := (p_J^K)^{-1}(A)$. Es gilt $B = B' = (p_J^K)^{-1}(A)$, also

$$P_K(B) = P_J(A)\,.$$

Für J und K beliebig setze $L := J \cup K$, also $J \subset L$, $K \subset L$. Also gibt es ein $C \in \mathcal{A}_L$ mit

$$p_L^{-1}(C) = p_J^{-1}(A) = p_K^{-1}(B)\,,$$

also $P_L(C) = P_J(A)$ und $P_L(C) = P_K(B)$, und somit $P_J(A) = P_K(B)$.

Somit ist durch $P_0\big(p_J^{-1}(A)\big) := P_J(A)$, $J \in \mathcal{H}$, $A \in \mathcal{A}_J$, eine Funktion P_0 auf $\mathcal{Z}$ definiert.

2. Schritt: $P_0 : \mathcal{Z} \to \mathbb{R}$ ist ein Inhalt, also $P_0 \geq 0$, $P_0(\emptyset) = 0$ und P_0 ist endlich additiv:

Seien $Y, Z \in \mathcal{Z}$, disjunkt, so existiert ein $J \in \mathcal{H}$ mit $Y = p_J^{-1}(A)$ und $Z = p_J^{-1}(B)$ für geeignete $A, B \in \mathcal{A}_J$. Mit Y und Z sind auch A und B disjunkt, $Y \cup Z = p_J^{-1}(A \cup B)$, also

$$P_0(Y \cup Z) = P_J(A \cup B) = P_J(A) + P_J(B) = P_0(Y) + P_0(Z)\,.$$

Wir zeigen noch die σ-Additivität von P_0, denn dann sind wir schon fertig: P ist dann die einzige Fortsetzung von P_0 zu einem Maß auf $\sigma(\mathcal{Z}) = \mathcal{A}$. P ist ein W-Maß, denn $\Omega = p_J^{-1}(\Omega_J)$, $J \in \mathcal{H}$, also ist Ω eine J-Zylindermenge und es gilt

$$P(\Omega) = P_0(\Omega) = P_J(\Omega_J) = 1\,.$$

3. Schritt: Zu $Z \in \mathcal{Z}$ und $J \in \mathcal{H}$ betrachte

$$Z^{\omega_J} := \Big\{\omega \in \Omega\,:\, \big(\omega_J, p_{I\setminus J}(\omega)\big) \in Z\Big\}\,.$$

Dies ist für jedes $\omega_J \in \Omega_J$ eine Zylindermenge:

Aus $Z = p_K^{-1}(A)$ für ein $K \in \mathcal{H}$, $A \in \mathcal{A}_K$, und ohne Einschränkung $J \subset K$ folgt

$$\begin{aligned} Z^{\omega_J} &= \big\{\omega'\,:\,(\omega_J,\, \omega'_{K\setminus J},\, \omega'_{I\setminus K}) \in Z\big\} \\ &= \big\{\omega'\,:\,(\omega_J,\, \omega_{K\setminus J}) \in A\big\} \\ &= \{\omega'\,:\,\omega'_{K\setminus J} \in A_{\omega_J}\} \end{aligned}$$

mit $A_{\omega_J} := \big\{\omega'_{K\setminus J}\,:\,(\omega_J,\, \omega_{K\setminus J}) \in A\big\}$, der ω_J-Schnitt von A in Ω_K. Also ist $Z^{\omega_J} = p_{K\setminus J}^{-1}(A_{\omega_J})$.

Es gilt

$$P_0(Z) = \int P_0(Z^{\omega_J})\, P_J(d\omega_J)\,.$$

Natürlich ist $A_{\omega_J} \in \mathcal{A}_{K\setminus J}$ und der Satz von Fubini liefert

$$P_0(Z) = P_K(A) = \int P_{K\setminus J}(A_{\omega_J})\, P_J(d\omega_J)\,.$$

Da $P_0(Z^{\omega_J}) = P_{K\setminus J}(A_{\omega_J})$, folgt also die Behauptung.

4. Schritt: Wir zeigen, dass P_0 $\emptyset$-stetig ist, d.h. dass für jede Folge $(B_n)_n$ von Ereignissen mit $B_n \downarrow \emptyset$ gilt: $\lim_{n\to\infty} P_0(B_n) = 0$. Daraus folgt die σ-Additivität nach Satz 1.19.

Sei $(Z_n)_n$ eine antitone Folge von Zylindermengen mit $\alpha := \inf_n P_0(Z_n) > 0$. Wir zeigen, dass $\bigcap_{n\ge 1} Z_n$ nicht leer sein kann.

Es gilt $Z_n = p_{J_n}^{-1}(A_n)$, $J_n \in \mathcal{H}$, $A_n \in \mathcal{A}_{J_n}$. Ohne Einschränkung kann $J_1 \subset J_2 \subset \ldots$ angenommen werden. Die Funktion

$$\omega_{J_1} \mapsto P_0(Z_n^{\omega_{J_1}})$$

ist $\mathcal{A}_{J_1}$-messbar (nach Lemma 8.6), also liegt

$$Q_n := \big\{\omega_{J_1} \in \Omega_{J_1}\,:\, P_0(Z_n^{\omega_{J_1}}) \ge \alpha/2\big\}$$

in $\mathcal{A}_{J_1}$, also folgt aus Schritt 3

$$\begin{aligned}\alpha \le P_0(Z_n) &= \Big(\int_{Q_n} + \int_{Q_n^c}\Big) P_0(Z_n^{\omega_{J_1}})\, P_{J_1}(d\omega_{J_1}) \\ &\le P_{J_1}(Q_n) + \frac{\alpha}{2}\,,\end{aligned}$$

also $P_{J_1}(Q_n) \ge \frac{\alpha}{2} > 0$. Mit $(Z_n)_n$ ist auch $(Q_n)_n$ antiton und P_{J_1} ist $\emptyset$-stetig, also kann $\bigcap_{n\ge1} Q_n$ nicht leer sein, es existiert also ein $\omega_{J_1} \in \bigcap_{n\ge1} Q_n$ mit

$$P_0(Z_n^{\omega_{J_1}}) \ge \frac{\alpha}{2} > 0 \quad \text{für alle } n \ge 1\,.$$

Ist $J_2 \ne J_1$, so liefert Schritt 3 die Existenz eines $\omega_{J_2\setminus J_1}$ mit

$$P_0\big((Z_n^{\omega_{J_1}})^{\omega_{J_2\setminus J_1}}\big) \ge \frac{\alpha}{4} \quad \text{für alle } n \ge 1\,.$$

Es gilt $\omega_{J_2} := (\omega_{J_1}, \omega_{J_2\setminus J_1}) \in \Omega_{J_2}$ und

$$(Z_n^{\omega_{J_1}})^{\omega_{J_2\setminus J_1}} = Z_n^{\omega_{J_2}}\,,$$

also

$$P_0(Z_n^{\omega_{J_2}}) \ge \frac{\alpha}{4} > 0 \quad \text{für alle } n \ge 1$$

und

$$\omega_{J_1} = p_{J_1}^{J_2}(\omega_{J_2})\,.$$

Ist $J_1 = J_2$, wählen wir $\omega_{J_2} = \omega_{J_1}$. Vollständige Induktion liefert: Zu jedem $k \in \mathbb{N}$ existiert ein $\omega_{J_k} \in \Omega_{J_k}$ mit

$$P_0(Z_n^{\omega_{J_k}}) \ge 2^{-k}\alpha > 0\,, \quad p_{J_k}^{J_{k+1}}(\omega_{J_{k+1}}) = \omega_{J_k}\,, \quad n, k \in \mathbb{N}\,.$$

Also existiert ein $\omega_0 \in \Omega$ mit $p_{J_k}(\omega_0) = \omega_{J_k}$ für alle $k \ge 1$ und $Z_n^{\omega_{J_n}} \ne \emptyset$, also existiert ein $\tilde{\omega}_n \in \Omega$ mit $\big(\omega_{J_n}, p_{I\setminus J_n}(\tilde{\omega}_n)\big) \in Z_n$.

In der J_n-Zylindermenge Z_n liegt dann auch der Punkt $\big(\omega_{J_n}, p_{I\setminus J_n}(\omega_0)\big) = \omega_0$. Also ist $\omega_0 \in Z_n$ für alle $n \ge 1$, d.h. $\bigcap_{n\ge1} Z_n \ne \emptyset$. □

8.20 Definition Das nach Satz 8.19 eindeutig bestimmte W-Maß P heißt das *Produktmaß der W-Maße* $(P_i)_{i\in I}$ und wird mit

$$\bigotimes_{i\in I} P_i$$

bezeichnet. Der W-Raum

$$\Big(\prod_{i\in I} \Omega_i\,, \bigotimes_{i\in I} \mathcal{A}_i\,, \bigotimes_{i\in I} P_i\Big)$$

heißt *Produkt der Familie* $\big((\Omega_i,\, \mathcal{A}_i,\, P_i)\big)_{i\in I}$ von W-Räumen und wird mit

$$\bigotimes_{i\in I}(\Omega_i,\, \mathcal{A}_i,\, P_i)$$

bezeichnet.

KAPITEL 9

Konvergenz von Zufallsvariablen und Verteilungen

Wir stellen vier verschiedene Konvergenzbegriffe für Folgen von Zufallsvariablen vor. Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $(X_n)_n$ eine Folge von Zufallsgrößen und X eine weitere Zufallsgröße auf $(\Omega, \mathcal{A}, P)$.

9.1 Definition Die Folge $(X_n)_n$ konvergiert P-fast sicher (P-f.s.) gegen die Zufallsgröße X, wenn gilt

$$P\big(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \text{ existiert und stimmt mit } X(\omega) \text{ überein}\}\big) = 1$$

oder kurz $P(X_n \to X) = 1$. Man schreibt dann $X_n \to X \; P-f.s.$ oder $X_n \to X$ $f.s.$, wenn über P kein Zweifel besteht.

$A = \{X_n \to X\}$ ist eine messbare Menge, denn

$$\begin{aligned} A &= \bigcap_{k\geq 1} \bigcup_{m\geq 1} \bigcap_{n\geq m} \big(\{|X_n - X| \leq 1/k, |X| < \infty\} \\ &\quad \cup \{X_n > k, X = \infty\} \cup \{X_n < -k, X = -\infty\}\big) \, . \end{aligned}$$

9.2 Lemma *Aus $X_n \to X$ f.s. und $X_n \to X'$ f.s. folgt $X = X'$ f.s.*

Beweis: In der Menge $\{X_n \to X\} \cap \{X_n \to X'\}$ gilt $X = X'$, also $P(X \neq X') \leq P(X_n \not\to X) + P(X_n \not\to X') = 0$. □

Wir betrachten nun ein notwendiges und hinreichendes Kriterium für die fast sichere Konvergenz:

9.3 Satz *$(X_n)_n$ konvergiert genau dann f.s. gegen einen reellen Limes X, wenn*

$$\lim_{m\to\infty} P\Big(\sup_{n\geq m} |X_n - X| > \varepsilon\Big) = 0 \;\; \textit{für alle } \varepsilon > 0.$$

Beweis: Da X reellwertig ist, konvergiert $(X_n)_n$ genau dann f.s. gegen X, wenn

$$\begin{aligned} A^c &= \bigcup_{k\geq 1} \bigcap_{m\geq 1} \bigcup_{n\geq m} \{|X_n - X| > 1/k\} \\ &= \bigcup_{k\geq 1} \bigcap_{m\geq 1} \Big\{ \sup_{n\geq m} |X_n - X| > 1/k \Big\} \end{aligned}$$

Wahrscheinlichkeit 0 hat, also

$$P\Big(\bigcap_{m\geq 1}\Big\{\sup_{n\geq m}|X_n - X| > 1/k\Big\}\Big) = \lim_{m\to\infty} P\Big(\sup_{n\geq m}|X_n - X| > 1/k\Big) = 0$$

für alle $k \geq 1$ gilt. □

9.4 Satz (Cauchy-Kriterium) *$X_n \to X$ f.s. für einen reellen Limes genau dann, wenn*

$$\lim_{m\to\infty} P\Big(\sup_{n\geq m}|X_n - X_m| > \varepsilon\Big) = 0$$

für alle $\varepsilon > 0$.

Beweis: „$\Rightarrow$ “:

$$\begin{aligned} P\Big(\sup_{n\geq m}|X_n - X_m| > \varepsilon\Big) &\leq P\Big(\sup_{n\geq m}|X_n - X| > \frac{\varepsilon}{2}\Big) + P\Big(|X_m - X| > \frac{\varepsilon}{2}\Big) \\ &\leq 2P\Big(\sup_{n\geq m}|X_n - X| > \frac{\varepsilon}{2}\Big). \end{aligned}$$

Nach Satz 9.3 konvergiert die rechte Seite der Ungleichungskette gegen Null für $m \to \infty$.

„$\Leftarrow$ “: Es sei $Y_n := \sup_{j,k\geq n}|X_k - X_j|$, $n \geq 1$. $(Y_n)_n$ ist eine monoton fallende Folge nichtnegativer Zufallsgrößen und es gilt

$$P\Big(\sup_{n\geq m} Y_n > \varepsilon\Big) = P(Y_m > \varepsilon) \leq 2P\Big(\sup_{n\geq m}|X_n - X_m| > \frac{\varepsilon}{2}\Big)\ .$$

Die rechte Seite konvergiert nach Voraussetzung gegen Null für $m \to \infty$. Nach Satz 9.3 folgt somit $Y_n \to 0$ f.s. Es sei $A := \{Y_n \to 0\}$, dann bildet $(X_n(\omega))_n$ für jedes $\omega \in A$ eine Cauchy-Folge in $\mathbb{R}$, hat also einen Limes $X(\omega)$. Für $\omega \in A^c$ setzen wir $X(\omega) = 0$, so dass $X_n \to X$ auf A, also fast sicher. □

Es sei an die $\mathcal{L}_p$-Konvergenz in Definition 7.11 erinnert. $(X_n)_n \subset \mathcal{L}_p(\Omega, \mathcal{A}, P)$, $p > 0$, konvergiert im p-ten Mittel gegen eine Zufallsgröße X, falls $X \in \mathcal{L}_p(\Omega, \mathcal{A}, P)$ und

$$\mathbb{E}(|X_n - X|^p) \to 0 \text{ für } n \to \infty\ .$$

Man spricht auch von $\mathcal{L}_p$-Konvergenz.

9.5 Definition $(X_n)_n$ *konvergiert in Wahrscheinlichkeit* (unter P) oder *P-stochastisch* gegen eine reelle Zufallsvariable X, wenn

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

für alle $\varepsilon > 0$ gilt.

Wir schreiben $X_n \xrightarrow{P} X$.

9.6 Lemma *Aus* $X_n \xrightarrow{P} X$ *und* $X_n \xrightarrow{P} X'$ *folgt* $X = X'$ *fast sicher.*

Beweis: $P(|X - X'| > \varepsilon) \leq P(|X_n - X| > \varepsilon/2) + P(|X_n - X'| > \varepsilon/2)$ für alle $n \geq 1$ und $\varepsilon > 0$. Da $\{X \neq X'\} = \bigcup_{k=1}^{\infty}\{|X - X'| \geq 1/k\}$, folgt die Behauptung. □

9.7 Definition $(X_n)_n$ konvergiert *schnell P-stochastisch* gegen X, wenn

$$\sum_{n\geq 1} P(|X_n - X| > \varepsilon) < \infty$$

für alle $\varepsilon > 0$ gilt.

9.8 Satz *Konvergiert $(X_n)_n$ schnell P-stochastisch gegen X, so auch fast sicher.*

Beweis: Es gilt

$$P(|X_n - X| > \varepsilon \text{ unendlich oft}) = P(\limsup\{|X_n - X| > \varepsilon\}) = 0$$

nach dem Lemma von Borel-Cantelli (1.21). Dies ist zu $\lim_{m\to\infty} P(\sup_{n\geq m} |X_n - X| > \varepsilon) = 0$ äquivalent, womit die Behauptung aus Satz 9.3 folgt. Die Äquivalenz sieht man so:

Ist $B_m := \bigcup_{n=m}^{\infty}\{|X_n - X| > \varepsilon\}$ und $B := \limsup_{n\to\infty}\{|X_n - X| > \varepsilon\}$, so gilt $B_m \searrow B$ und $\lim_{m\to\infty} P(B_m) = P(B)$ und $B_m = \{\sup_{n\geq m} |X_n - X| > \varepsilon\}$. □

9.9 Satz *Fast sichere Konvergenz impliziert Konvergenz in Wahrscheinlichkeit. Konvergenz im p-ten Mittel, $p \geq 1$, impliziert Konvergenz in Wahrscheinlichkeit.*

Beweis: Die Ungleichung $P(|X_n - X| > \varepsilon) \leq P(\sup_{k\geq n} |X_k - X| > \varepsilon)$ liefert mit Satz 9.3 die erste Aussage, die zweite folgt aus der Markov-Ungleichung 7.17. □

Die anderen denkbaren Implikationen sind nicht richtig, wie die folgenden Beispiele belegen:

9.10 Beispiele Es sei $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B} \cap [0,1], \lambda|_{[0,1]})$.

(a) Sei $X_n = n^{1/p} 1_{[0,1/n]}$, $p > 0$. Dann gilt $X_n \to 0$ f.s. und in W.keit, aber $\mathbb{E}(|X_n|^p) = 1$ für alle $n \in \mathbb{N}$, also konvergiert $(X_n)_n$ nicht im p-ten Mittel gegen Null.

(b) Für jede natürliche Zahl n gibt es genau ein Paar ganzer Zahlen $m \geq 0$, $k \geq 0$ mit $n = 2^m + k$ und für $k < 2^m$. Wir setzen $X_n = 1_{[k2^{-m},(k+1)2^{-m}[}$. $(X_n(\omega))_n$ konvergiert für kein $\omega \in [0,1]$. Zu $\omega \in \Omega$ und $m = 0, 1, \ldots$ existiert genau ein $k \in \{0, \ldots, 2^m - 1\}$ mit $\omega \in [k2^{-m}, (k+1)2^{-m}[$. Im Fall $k < 2^m - 1$ ist $\omega \notin [(k+1)2^{-m}, (k+2)2^{-m}[$ und im Fall $k = 2^m - 1$ und $k \geq 1$ ist $\omega \notin [0, 2^{-(m+1)}[$. Aber es ist $P(|X_n| > \varepsilon) \leq 2^{-m}$ für alle $\varepsilon > 0$ und $\mathbb{E}(|X_n|^p) = 2^{-m}$ für $p > 0$.

Unter Zusatzbedingungen impliziert die fast sichere Konvergenz die Konvergenz im p-ten Mittel:

9.11 Satz *Es sei $(X_n)_n$ eine Folge von Zufallsgrößen, die f.s. gegen X konvergiert. Gilt $|X_n| \leq Y$ fast sicher für ein $Y \in \mathcal{L}_p$, $p > 0$, so konvergiert X_n gegen X im p-ten Mittel.*

Beweis: Es gilt

$$|X_n - X|^p < (|X_n| + |X|)^p \leq (2Y)^p = 2^p Y^p \in \mathcal{L}_1.$$

Weiter ist $|X_n - X|^p \longrightarrow 0$ fast sicher. Nach dem Satz von der dominierten Konvergenz folgt $\mathbb{E}(|X_n - X|^p) \to 0$. □

Aus der Konvergenz in W.keit folgt nicht fast sichere Konvergenz (siehe Beispiel 9.10), aber es gilt:

9.12 Satz *Es sei $(X_n)_n$ eine Folge, die in W.keit gegen X konvergiert, so existiert eine Teilfolge $(X_{n_k})_k$ mit $\lim_{k\to\infty} X_{n_k} = X$ fast sicher.*

Beweis: Zu jedem $k \in \mathbb{N}$ existiert ein $n_k \in \mathbb{N}$ mit $P(|X_{n_k} - X| \geq 1/k) \leq 1/k^2$. Wir können $n_{k+1} > n_k$ für alle $k \in \mathbb{N}$ annehmen. Da $\sum_{k\geq 1} 1/k^2 < \infty$ folgt die Behauptung aus Satz 9.8. □

Alle drei bisherigen Konvergenztypen (fast sicher, im p-ten Mittel und in W.keit) sind vollständig, das heißt, dass jede Cauchy-Folge konvergiert. Für fast sichere Konvergenz folgt dies unmittelbar aus der Vollständigkeit von $\mathbb{R}$, für Konvergenz im p-ten Mittel hatten wir es in Kapitel 7 (Satz von Riesz-Fischer) gesehen, für Konvergenz in W.keit ist es der folgende Satz:

9.13 Satz *Es sei* $(X_n)_n$ *eine Folge von Zufallsgrößen mit*

$$\lim_{n,m\to\infty} P(|X_n - X_m| \geq \varepsilon) = 0$$

für alle $\varepsilon > 0$. *Dann existiert eine Zufallsgröße* X *mit* $X_n \to X$ *in W.keit.*

Beweis: Wähle wie im Beweis von Satz 9.12 eine Teilfolge $(n_k)_k$ mit

$$P(|X_{n_k} - X_{n_{(k+1)}}| \geq 1/k) \leq 1/k^2 \ .$$

Nach dem Lemma von BOREL-CANTELLI folgt

$$P\big(\limsup_{k\to\infty}\{|X_{n_k} - X_{n_{(k+1)}}| \geq 1/k\}\big) = 0 \ .$$

Für $\omega \notin \limsup_{k\to\infty}\{|X_{n_k} - X_{n_{(k+1)}}| \geq 1/k\}$ ist $(X_{n_k}(\omega))_k$ also eine CAUCHY-Folge in $\mathbb{R}$, also konvergiert $(X_{n_k})_k$ f.s. gegen ein X, also in W.keit. Für $\varepsilon > 0$ gilt

$$P(|X_m - X| \geq \varepsilon) \leq P(|X_m - X_{n_k}| > \varepsilon/2) + P(|X_{n_k} - X| > \varepsilon/2)$$

für alle m und k. Wähle k als die kleinste Zahl mit $n_k \geq m$, so folgt

$$\lim_{m\to\infty} P(|X_m - X| \geq \varepsilon) = 0$$

für jedes $\varepsilon > 0$. □

Die Verteilung einer Zufallsvariablen spielt eine zentrale Rolle. Implizieren die diskutierten Arten der Konvergenz von $(X_n)_n$ gegen X eine Konvergenz der Folge der Verteilungen P^{X_n} gegen die Verteilung P^X von X? Was soll dabei eine Konvergenz von Verteilungen genau bedeuten?

9.14 Definition Eine Folge $(\mu_n)_n$ von W-Maßen auf $\mathcal{B}^d$ heißt *schwach konvergent* gegen ein W-Maß μ auf $\mathcal{B}^d$, wenn

$$\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu$$

für alle Funktionen $f \in C_b(\mathbb{R}^d)$ gilt (wobei $C_b(\mathbb{R}^d)$ den Vektorraum aller beschränkten, stetigen, reellen Funktionen auf $\mathbb{R}^d$ bezeichnet). Man schreibt $\lim_{n\to\infty} \mu_n = \mu$.

Sind X und $(X_n)_n$ $\mathbb{R}^d$-wertige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ und konvergiert $(P^{X_n})_n$ schwach gegen P^X bzw. allgemeiner gegen ein W-Maß ν auf $\mathcal{B}^d$, so nennt man die Folge $(X_n)_n$ *konvergent in Verteilung* gegen X bzw. gegen ν.

9.15 Bemerkungen

(a) Warum ist nicht $\mu_n(A) \to \mu(A)$ für alle $A \in \mathcal{B}^d$ der geeignete Konvergenzbegriff? Es sei μ_n die Verteilung einer binomialverteilten Zufallsgröße X_n zu den Parametern n und p. Dann existiert zu jedem n eine endliche Menge A_n mit $P\Big(\frac{X_n - np}{\sqrt{np(1-p)}} \in A_n\Big) = 1$. Für $A := \bigcup_{n=1}^\infty A_n$ gilt dann $P\Big(\frac{X_n - np}{\sqrt{np(1-p)}} \in A\Big) = 1$ für alle $n \in \mathbb{N}$, aber $\nu_{0,1}(A) = 0$ und nach dem Satz von DE MOIVRE und LAPLACE erwarten wir $\nu_{0,1}$ als „Limesverteilung“!

(b) Sei $(x_n)_n$ eine Folge reeller Zahlen. Sie konvergiert genau dann gegen $x_0 \in \mathbb{R}$, wenn die Folge $(\delta_{x_n})_n$ schwach gegen δ_{x_0} konvergiert. Aus $\lim_{n\to\infty} x_n = x_0$ folgt sofort $\lim_{n\to\infty} \delta_{x_n} = \delta_{x_0}$. Für die Rückrichtung sei zu $\varepsilon > 0$

$$f(x) := \max(0, 1 - 1/\varepsilon|x - x_0|) \ ,$$

also ist $f \in C_b(\mathbb{R})$. Es gilt $\{f > 0\} = (x_0 - \varepsilon, x_0 + \varepsilon)$ und $\lim_{n\to\infty} f(x_n) = f(x_0) = 1$. Also gilt $|x_n - x_0| < \varepsilon$ für schließlich alle $n \in \mathbb{N}$.

(c) $(\Omega, \mathcal{A}, P)$ sei gewählt wie in Beispiel 9.10, $X_n := 1_{[1/2,1)}$, $X := 1_{[0,1/2]}$. Dann ist $P^{X_n} = P^X = (1/2)(\delta_0 + \delta_1)$. $(X_n)_n$ konvergiert also in Verteilung sowohl gegen X als auch gegen X_1. Aber $|X(\omega) - X_1(\omega)| = 1$ für $\omega \in \Omega$, also $X(\omega) \neq X_1(\omega)$ für alle $\omega \in \Omega$ und $P(|X - X_1| \geq 1) = 1$ für jedes $n \in \mathbb{N}$. Somit ist bei Konvergenz in Verteilung der Limes nicht fast sicher eindeutig bestimmt und aus der Konvergenz in Verteilung folgt im Allgemeinen nicht die stochastische Konvergenz.

Es gilt aber

9.16 Satz *Eine Folge $(X_n)_n$ reeller Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ konvergiere stochastisch gegen eine reelle Zufallsgröße X auf Ω. Dann konvergiert $(X_n)_n$ in Verteilung gegen X. Ist X fast sicher konstant, also P^X ein Dirac-Maß, so gilt hiervon auch die Umkehrung.*

Wir fassen vor dem Beweis von Satz 9.16 die Zusammenhänge zwischen den Konvergenzbegriffen in einem Schema zusammen:

$$\begin{array}{ccc}
 & & \mathcal{L}^p\text{-Konvergenz} \\
 & & \Downarrow \\
\text{fast sichere Konvergenz} & & \mathcal{L}^1\text{-Konvergenz} \\
\searrow & & \swarrow \\
 & \text{stochastische Konvergenz} & \\
 & \Downarrow & \\
 & \text{schwache Konvergenz der Verteilungen} & \\
 & \text{(Konvergenz in Verteilung)} &
\end{array}$$

Beweis: Sei $f \in C_b(\mathbb{R})$ zunächst gleichmäßig stetig auf $\mathbb{R}$. Zu $\varepsilon > 0$ existiert also ein $\delta > 0$, so dass für $|x-y| < \delta$ folgt $|f(x) - f(y)| < \varepsilon$, $x, y \in \mathbb{R}$. Es sei $A_n := \{|X_n - X| \geq \delta\}$, $n \in \mathbb{N}$. Es gilt $|f \circ X_n - f \circ X| \leq |f \circ X_n| + |f \circ X| \leq 2\|f\|_\infty$ (Supremums-Norm). Es folgt

$$\begin{aligned}
\left| \int f \, dP^{X_n} - \int f \, dP^X \right| &= |\mathbb{E}(f \circ X_n - f \circ X)| \\
&\leq \mathbb{E}|f \circ X_n - f \circ X| \\
&= \int_{A_n} |f \circ X_n - f \circ X| \, dP + \int_{A_n^c} |f \circ X_n - f \circ X| \, dP \\
&\leq 2\|f\|_\infty P(A_n) + \varepsilon P({A_n}^c) \\
&\leq 2\|f\|_\infty P(A_n) + \varepsilon \; .
\end{aligned}$$

Wegen der stochastischen Konvergenz ist $\lim_{n\to\infty} P(A_n) = 0$, also folgt $\lim_{n\to\infty} \int f \, dP^{X_n} = \int f \, dP^X$ für diese Klasse von Abbildungen.

Es sei f nun beliebig (in $C_b(\mathbb{R})$). Wähle $I_n := [-n, n] \nearrow \mathbb{R}$, also $P^X(I_n) \nearrow 1$. Zu jedem $\varepsilon > 0$ existiert daher ein $n_0 \in \mathbb{N}$ mit $1 - P^X(I_{n_0}) = P^X(\mathbb{R} \setminus I_{n_0}) < \varepsilon$. Eine Funktion $u_\varepsilon \in C_b(\mathbb{R})$ sei wie folgt definiert: auf I_{n_0} sei sie gleich 1, auf $[n_0, n_0+1]$ und $[-n_0-1, -n_0]$ affin-linear, $u_\varepsilon(n_0+1) = u_\varepsilon(-n_0-1) = 0$ und auf $I_{n_0+1}^c$ sei sie Null (siehe Abbildung 9.1).

Wir betrachten nun $f' := u_\varepsilon f$. Die Funktionen u_ε und f' sind auf $I_{n_0+1}^c$ Null und daher auf $\mathbb{R}$ gleichmäßig stetig. Damit folgt

$$\lim_{n\to\infty} \int f' \, dP^{X_n} = \int f' \, dP^X \qquad \lim_{n\to\infty} \int u_\varepsilon \, dP^{X_n} = \int u_\varepsilon \, dP^X$$

nach bereits Gezeigtem. Also folgt auch

$$\lim_{n\to\infty} \int (1-u_\varepsilon) \, dP^{X_n} = \int (1-u_\varepsilon) \, dP^X \; . \tag{9.1}$$

Nun gilt

$$\begin{aligned}
&\left| \int f \, dP^{X_n} - \int f \, dP^X \right| \\
&\leq \int |f - f'| \, dP^{X_n} + \left| \int f' \, dP^{X_n} - \int f' \, dP^X \right| + \int |f' - f| \, dP^X
\end{aligned} \tag{9.2}$$

und

$$\int |f - f'| \, dP^X = \int |f|(1-u_\varepsilon) \, dP^X \leq \|f\|_\infty \varepsilon \; ,$$

denn $\int (1-u_\varepsilon) \, dP^X \leq P^X(\mathbb{R} \setminus I_{n_0}) < \varepsilon$. Weiter ist dann $\int (1-u_\varepsilon) \, dP^{X_n} < \varepsilon$ nach (9.1) für schließlich alle n, etwa $n \geq n_1$. Also folgt auch

$$\int |f - f'| \, dP^{X_n} \leq \|f\|_\infty \varepsilon$$

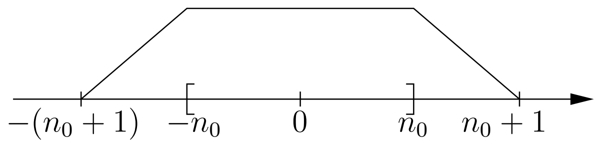


Abb. 9.1:

für alle $n \geq n_1$. Also ist die rechte Seite von (9.2) für n hinreichend groß kleiner $2\|f\|_\infty \varepsilon + \varepsilon$, was zu zeigen war.

Ist umgekehrt $X = \eta$ P-fast sicher, also $P^X = \delta_\eta$, so wähle zu $(\eta-\varepsilon, \eta+\varepsilon)$ mit $\varepsilon > 0$ eine stückweise affin-lineare Funktion $f \in C_b(\mathbb{R})$ mit $f \leq 1_{(\eta-\varepsilon,\eta+\varepsilon)}$ und $f(\eta) = 1$. Dann ist

$$\int f\, dP^{X_n} \leq P^{X_n}\big((\eta-\varepsilon, \eta+\varepsilon)\big) = P\big(X_n \in (\eta-\varepsilon, \eta+\varepsilon)\big) \leq 1$$

und nach Voraussetzung $\lim_{n\to\infty} \int f\, dP^{X_n} = f(\eta) = 1$, also folgt

$$\lim_{n\to\infty} P\big(X_n \in (\eta-\varepsilon, \eta+\varepsilon)\big) = 1 \ .$$

Nun ist $\{X_n \in (\eta-\varepsilon, \eta+\varepsilon)\} = \{|X_n - \eta| < \varepsilon\}$, also $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$ für alle $\varepsilon > 0$, womit die stochastische Konvergenz gezeigt ist. □

Der Beweis des obigen Satzes zeigt insbesondere, dass $(P^{X_n})_n$ schwach gegen P^X konvergiert genau dann, wenn

$$\lim_{n\to\infty} \int f\, dP^{X_n} = \int f\, dP^X$$

für alle *gleichmäßig stetigen* und beschränkten $f : \mathbb{R} \to \mathbb{R}$ gilt.

KAPITEL 10

Unabhängigkeit

In Kapitel 8 haben wir das Produktwahrscheinlichkeitsmaß auf unendlichen Produkträumen konstruiert, um ein Zufallsexperiment mit unendlich vielen Einzelexperimenten zu beschreiben. Wenn sich die Ausgänge der Einzelexperimente nicht gegenseitig beeinflussen, spricht man von „stochastisch unabhängigen" Experimenten. Wir wollen diesen Begriff präzisieren. Er basiert letztendlich auf dem Begriff des Produktmaßes.

Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Wir setzen im Folgenden voraus, dass Familien von Teilmengen von Ω stets Ω enthalten.

10.1 Definition

(a) Teilmengen $\mathcal{E}_1, \ldots, \mathcal{E}_n$ von $\mathcal{A}$ mit $\Omega \in \mathcal{E}_i$ heißen *unabhängig*, wenn für $A_i \in \mathcal{E}_i$, $1 \leq i \leq n$, gilt:

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)\,.$$

(b) Es sei I eine Indexmenge und $\mathcal{E}_i$ für $i \in I$ seien Teilmengen von $\mathcal{A}$. Sie heißen *unabhängig*, wenn je endlich viele unabhängig sind.

(c) Ereignisse A_i für $i \in I$ heißen *unabhängig*, wenn die Mengensysteme $\{A_i, \Omega\}$, $i \in I$, unabhängig sind.

10.2 Bemerkung Die Voraussetzung, dass die Mengensysteme stets Ω enthalten, dient der bequemen Notation. Sie hat nämlich zur Folge, dass für unabhängige Mengensysteme $\mathcal{E}_1, \ldots, \mathcal{E}_n$ auch stets

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}) \tag{10.1}$$

für $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ und $A_{i_j} \in E_{i_j}$ ist. Setzt man $\Omega \in \mathcal{E}_i$ nicht voraus, so muss man (10.1) als Definition verwenden.

10.3 Lemma *Sind die $\mathcal{E}_i$ für $i \in I$ unabhängig und gilt $\mathcal{D}_i \subset \mathcal{E}_i$ für $i \in I$, so sind die $\mathcal{D}_i$ für $i \in I$ unabhängig. Ist $\mathcal{D}$ unabhängig von $\mathcal{E}_i$ für $i \in I$, so ist $\mathcal{D}$ unabhängig von $\bigcup_{i \in I} \mathcal{E}_i$.*

Beweis: Der erste Teil ist klar. Für $A \in \mathcal{D}$ und $B \in \bigcup_{i \in I} \mathcal{E}_i$ existiert ein $i \in I$ mit $B \in \mathcal{E}_i$, also ist

$$P(A \cap B) = P(A)\, P(B)\,. \qquad \square$$

10.4 Beispiele

(a) Es seien $(\Omega_i,\, \mathcal{A}_i,\, P_i)$, $i = 1, \dots, n$, endlich viele W-Räume, $(\Omega,\, \mathcal{A},\, P)$ der Produktraum und $\tilde{A}_i \in \mathcal{A}_i$, $i = 1, \dots, n$. Dann sind

$$\begin{aligned} A_1 &:= \tilde{A}_1 \times \Omega_2 \times \cdots \times \Omega_n\,, \\ A_2 &:= \Omega_1 \times \tilde{A}_2 \times \Omega_3 \times \cdots \times \Omega_n\,, \\ \dots & \quad \dots \\ A_n &:= \Omega_1 \times \cdots \times \Omega_{n-1} \times \tilde{A}_n \end{aligned}$$

unabhängig.

(b) Es seien $\Omega := [0,\, 1)$, $\mathcal{A} := \Omega \cap \mathcal{B}^1$, $P := \lambda^1|_\Omega$. Für $n \in \mathbb{N}$ sei

$$A_n := \left[0,\, \frac{1}{2^n}\right) \cup \left[\frac{2}{2^n},\, \frac{3}{2^n}\right) \cup \ldots \cup \left[\frac{2^n-2}{2^n},\, \frac{2^n-1}{2^n}\right)\,.$$

Die $(A_n)_n$ sind unabhängig, denn $P(A_n) = \frac{1}{2}$, $n \in \mathbb{N}$, und

$$\begin{aligned} P(A_{i_1} \cap \cdots \cap A_{i_n}) &= \frac{1}{2} P(A_{i_1} \cap \cdots \cap A_{i_{n-1}}) \\ &= P(A_{i_1}) \cdots P(A_{i_n}) \end{aligned}$$

für je endlich viele paarweise verschiedene Zahlen $i_1, \dots, i_n \in \mathbb{N}$. Die Menge A_n ist die Menge aller $x \in [0,\, 1)$ mit $\varepsilon_n = 0$ in der eindeutigen dyadischen Entwicklung

$$x = \sum_{k=1}^{\infty} \varepsilon_k 2^{-k}$$

mit $\varepsilon_k = 0$ oder 1 und nicht $\varepsilon_k = 1$ für schließlich alle k.

Wir diskutieren nun Möglichkeiten, Unabhängigkeitsaussagen von Mengensystemen auf größere Mengensysteme hochzuziehen:

10.5 Satz *Es seien $\mathcal{D}_i$ für $i \in I$ unabhängige Teilmengen von $\mathcal{A}$ mit $\Omega \in \mathcal{D}_i$. Sind die $\mathcal{D}_i$ durchschnittstabil, so sind die $\sigma(\mathcal{D}_i)$ für $i \in I$ unabhängig.*

Beweis: Ohne Einschränkung sei I endlich, etwa $I = \{1, \dots, n\}$. Wir zeigen

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n) \tag{10.2}$$

für $A_i \in \sigma(\mathcal{D}_i)$. Für $0 \le k \le n$ sei L_k die folgende Aussage:

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)\,, \qquad \begin{aligned} &\forall A_i \in \sigma(\mathcal{D}_i) \text{ für } i \le k\,, \\ &\forall A_i \in \mathcal{D}_i \text{ für } i > k\,. \end{aligned}$$

L_0 gilt, da die $\mathcal{D}_i$ unabhängig sind. Wir zeigen

$$L_k \Rightarrow L_{k+1} \text{ für } 0 \leq k \leq n-1\,.$$

Betrachte das Mengensystem $\mathcal{A}_{k+1}$ bestehend aus den Mengen $A_{k+1} \in \sigma(\mathcal{D}_{k+1})$, die die Eigenschaft haben, dass die Gleichung (10.2) $\forall A_1 \in \sigma(\mathcal{D}_1), \ldots, \forall A_k \in \sigma(\mathcal{D}_k)$, $\forall A_{k+2} \in \mathcal{D}_{k+2}, \ldots, \forall A_n \in \mathcal{D}_n$ gilt.

Aus L_k folgt $\mathcal{A}_{k+1} \supset \mathcal{D}_{k+1}$. Wir zeigen, dass $\mathcal{A}_{k+1}$ ein DYNKIN-System ist.

(a) $\Omega \in \mathcal{A}_{k+1}$ gilt, denn $\Omega \in \mathcal{D}_{k+1}$.

(b) Für $D \in \mathcal{A}_{k+1}$ gilt

$$\begin{aligned}
&P\Big(\bigcap_{j=1}^{k} A_j \cap D^c \cap \bigcap_{j=k+2}^{n} A_j\Big) \\
&\qquad = P\big(\bigcap_{j=1}^{k} A_j \cap \bigcap_{j=k+2}^{n} A_j\Big) - P\big(\bigcap_{j=1}^{k} A_j \cap D \cap \bigcap_{j=k+2}^{n} A_j\Big) \\
&\qquad = \prod_{j,\, j\neq k+1} P(A_j) - P(D) \prod_{j,\, j\neq k+1} P(A_j) \\
&\qquad = \prod_{j,\, j\neq k+1} P(A_j)\, P(D^c)
\end{aligned}$$

für alle A_i gemäß den obigen Bedingungen, also $D^c \in \mathcal{A}_{k+1}$.

(c) Für paarweise disjunkte $D_i \in \mathcal{A}_{k+1}$, $i \in \mathbb{N}$, folgt mittels der σ-Additivität von P

$$\bigcup_{i=1}^{\infty} D_i \in \mathcal{A}_{k+1}\,.$$

Nun folgt aus Satz 1.11

$$\mathcal{A}_{k+1} = \sigma(\mathcal{D}_{k+1})\,,$$

was aber heißt, dass L_{k+1} gilt. □

10.6 Bemerkung Da das Mengensystem $\{A, \Omega\}$, $A \in \mathcal{A}$, durchschnittstabil ist, folgt: Sind A_i für $i \in I$ unabhängige Ereignisse, so sind die σ-Algebren $\{\emptyset, A_i, A_i^c, \Omega\}$ unabhängig, insbesondere auch die Komplemente A_i^c.

10.7 Korollar (Blockbildung) *Es seien $\mathcal{D}_i \subset \mathcal{A}$ für $i \in I$ unabhängig und durchschnittstabil. Es sei $(I_k)_{k\in K}$ eine Familie von paarweise disjunkten Teilmengen von I. Dann sind die*

$$\sigma\Big(\bigcup_{j\in I_k} \mathcal{D}_j\Big)$$

für $k \in K$ unabhängig.

Beweis: Für $k \in K$ sei $\hat{\mathcal{D}}_k$ die Familie der endlichen Durchschnitte von Elementen aus $\mathcal{D}_j$ für $j \in I_k$. $\hat{\mathcal{D}}_k$ ist durchschnittstabil, und da die $\mathcal{D}_j$ durchschnittstabil sind, hat jedes Element aus $\hat{\mathcal{D}}_k$ die Gestalt $A_{j_1} \cap \cdots \cap A_{j_n}$ mit $n \in \mathbb{N}$, $A_j \in \mathcal{D}_j$ und $j_1, \ldots, j_n \in I_k$ verschieden. Daraus folgt, dass die $\hat{\mathcal{D}}_k$ für $k \in K$ unabhängig sind. Da $\hat{\mathcal{D}}_k \supset \mathcal{D}_j$ für alle $j \in I_k$, gilt

$$\sigma\Big(\bigcup_{j \in I_k} \mathcal{D}_j\Big) \subset \sigma\Big(\hat{\mathcal{D}}_k\Big) .$$

Also folgt die Behauptung aus Satz 10.5. □

10.8 Definition Es sei $(\mathcal{A}_n)_n$ eine Folge von σ-Algebren von Ereignissen aus $\mathcal{A}$ und

$$\mathcal{T}_n := \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{A}_m\Big) .$$

Dann heißt

$$\mathcal{T}_\infty := \bigcap_{n=1}^{\infty} \mathcal{T}_n$$

die σ-Algebra der *terminalen Ereignisse* der Folge $(\mathcal{A}_n)_n$.

10.9 Satz (Null-Eins-Gesetz von Kolmogorov) *Es sei $(\mathcal{A}_n)_n$ eine unabhängige Folge von σ-Algebren $\mathcal{A}_n \subset \mathcal{A}$. Dann gilt*

$$P(A) \in \{0, 1\}$$

für $A \in \mathcal{T}_\infty$.

Beweis: Nach Korollar 10.7 ist $\mathcal{T}_{n+1}$ unabhängig von $\sigma\big(\bigcup_{m=1}^n \mathcal{A}_m\big)$ und somit ist $\mathcal{T}_\infty$ unabhängig von $\sigma\big(\bigcup_{m=1}^n \mathcal{A}_m\big)$ (Lemma 10.3) für alle $n \in \mathbb{N}$. Dann ist $\mathcal{T}_\infty$ unabhängig von

$$\bigcup_{n=1}^{\infty} \sigma\Big(\bigcup_{m=1}^{n} \mathcal{A}_m\Big)$$

nach Lemma 10.3. Eine Vereinigung von aufsteigenden Mengen ist durchschnittstabil, also folgt mit Satz 10.5, dass $\mathcal{T}_\infty$ unabhängig ist von

$$\sigma\Big(\bigcup_{n=1}^{\infty} \sigma(\bigcup_{m=1}^{n} \mathcal{A}_m)\Big) = \sigma\Big(\bigcup_{m=1}^{\infty} \mathcal{A}_m\Big) .$$

Natürlich ist $\mathcal{T}_n \subset \sigma\big(\bigcup_{m=1}^\infty \mathcal{A}_m\big)$ für alle $n \in \mathbb{N}$, also auch

$$\mathcal{T}_\infty \subset \sigma\Big(\bigcup_{m=1}^\infty \mathcal{A}_m\Big),$$

also ist nach Lemma 10.3 $\mathcal{T}_\infty$ unabhängig zu sich selbst! Das heißt für $A \in \mathcal{T}_\infty$ gilt

$$P(A) = P(A \cap A) = P(A)^2$$

also $P(A) \in \{0, 1\}$. □

10.10 Korollar (Null-Eins-Gesetz von Borel) *Für jede unabhängige Folge $(A_n)_n$ von Ereignissen aus $\mathcal{A}$ gilt*

$$P(\limsup_{n\to\infty} A_n) = 0 \textit{ oder } = 1\,.$$

Beweis: Nach Satz 10.5 ist $\mathcal{A}_n := \sigma\big(\{A_n, \Omega\}\big)$ eine unabhängige Folge. Es gilt $Q_n := \bigcup_{m=n}^\infty A_m \in \mathcal{T}_n$, sogar $Q_m \in \mathcal{T}_n$ für jedes $m \geq n$, $m \in \mathbb{N}$.

$$\limsup_{n\to\infty} A_n = \bigcap_{k=1}^\infty Q_k = \bigcap_{k=j}^\infty Q_k \in \mathcal{T}_j$$

für alle $j \in \mathbb{N}$, da $(Q_n)_n$ antiton ist, also ist $\limsup A_n \in \mathcal{T}_\infty$. □

Aus dem Lemma von Borel-Cantelli, 1.21, wissen wir

$$\sum_{n\geq 1} P(A_n) < \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 0\,.$$

Die Divergenz von $\sum P(A_n)$ führt im Allgemeinen nicht zum Schluss $P(\limsup A_n) = 1$. Wählt man nämlich ein $A_0 \in \mathcal{A}$ mit $0 < P(A_0) < 1$ und $(A_n)_n$ als konstante Folge $A_0, A_0, \ldots$, dann divergiert $\sum P(A_n)$, aber $P(\limsup A_n) = P(A_0) < 1$.

Nimmt man unabhängige $(A_n)_n$, so gilt die Umkehrung von 1.21, was auch von Borel und Cantelli bewiesen wurde. Es genügt, paarweise Unabhängigkeit zu fordern, was auf Erdős und Rényi zurückgeht:

10.11 Satz (von Borel-Cantelli, Erdős-Rényi) *Sei $(A_n)_n$ eine Folge von Ereignissen in einem W-Raum $(\Omega, \mathcal{A}, P)$. Dann gilt:*

$$\sum_{n\geq 1} P(A_n) < \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 0\,.$$

Sind die Ereignisse wenigstens paarweise unabhängig, so gilt

$$\sum_{n\geq 1} P(A_n) = \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 1\,.$$

Beweis: Es sei $A := \limsup A_n$. Der erste Teil ist Lemma 1.21. Den zweiten Teil beweisen wir zunächst für unabhängige Ereignisse, weil der Beweis klassisch und kurz ist.

Mit der Stetigkeit von W-Maßen folgt

$$P(A^c) = P(\liminf_{n\to\infty} A_n^c) = \lim_{n\to\infty} P(\bigcap_{k\geq n} A_k^c)\,.$$

Die $(A_n^c)_n$ sind unabhängig und $\sum_{k\geq n} P(A_k) = \infty$ für alle $n \in \mathbb{N}$, also

$$P(A^c) = \lim_{n\to\infty} \prod_{k\geq n} P(A_k^c) = \lim_{n\to\infty} \exp\Big(\sum_{k\geq n} \log\big(1 - P(A_k)\big)\Big)\,.$$

Für $x \in [0,\, 1]$ gilt $\log(1-x) \leq -x$, also

$$P(A^c) \leq \lim_{n\to\infty} \exp\Big(-\sum_{k\geq n} P(A_k)\Big) = 0\,.$$

Im Fall paarweise unabhängiger Ereignisse $(A_n)_n$ setzen wir

$$I_n := 1_{A_n}\,, \quad S_n = \sum_{j=1}^{n} I_j \quad \text{und} \quad S := \sum_{n=1}^{\infty} I_n\,.$$

Die I_n sind nach Voraussetzung paarweise unkorreliert. Weiter ist $I_n^2 = I_n$. Also ist

$$\begin{aligned}\mathrm{Var}(S_n) &= \sum_{j=1}^{n} \mathrm{Var}(I_j) = \sum_{j=1}^{n} \big(\mathbb{E}(I_j^2) - \mathbb{E}(I_j)^2\big) \\ &= \mathbb{E}(S_n) - \sum_{j=1}^{n} \mathbb{E}(I_j)^2 \leq \mathbb{E}(S_n)\,.\end{aligned}$$

Die Voraussetzung besagt $\sum_{n=1}^{\infty} \mathbb{E}(I_n) = +\infty$ und daher folgt wegen $S_n \uparrow S$

$$\lim_{n\to\infty} \mathbb{E}(S_n) = \mathbb{E}(S) = +\infty\,. \tag{10.3}$$

Ein Element $\omega \in \Omega$ liegt genau dann in A, also in A_n für unendlich viele n, wenn $S(\omega) = \infty$ ist. Zu zeigen ist also $P(S = +\infty) = 1$.

Nach Tschebyschev ist

$$P\big(|S_n - \mathbb{E}(S_n)| \leq \eta\big) \geq 1 - \frac{\mathrm{Var}(S_n)}{\eta^2}$$

für $\eta > 0$. Mit (10.3) kann $\mathbb{E}(S_n) > 0$ für alle $n \in \mathbb{N}$ angenommen werden. Es folgt

$$\begin{aligned}P\big(S_n \geq \tfrac{1}{2}\mathbb{E}(S_n)\big) &\geq P\big(|S_n - \mathbb{E}(S_n)| \leq \tfrac{1}{2}\mathbb{E}(S_n)\big) \\ &\geq 1 - 4\frac{\mathrm{Var}(S_n)}{\mathbb{E}(S_n)^2}\,.\end{aligned}$$

Es ist $\lim_{n\to\infty} \mathrm{Var}(S_n)/\mathbb{E}(S_n)^2 = 0$, und somit

$$P\big(S_n \geq \frac{1}{2}\mathbb{E}(S_n)\big) \geq 1 - \varepsilon$$

für jedes $\varepsilon > 0$ und schließlich alle n.

Da $S_n \leq S$, folgt

$$P\big(S \geq \frac{1}{2}\mathbb{E}(S_n)\big) \geq P\big(S_n \geq \frac{1}{2}\mathbb{E}(S_n)\big) \geq 1 - \varepsilon$$

für schließlich alle n. Nach (10.3) gilt $\mathbb{E}(S_n) \uparrow \mathbb{E}(S) = +\infty$, also

$$P(S = +\infty) \geq 1 - \varepsilon$$

für alle $\varepsilon > 0$, also $P(S = \infty) = 1$. $\square$

Die Ereignisse $(A_i)_{i\in I}$ sind genau dann unabhängig, wenn die $\mathcal{A}_i = \{\Omega, \emptyset, A_i, A_i^c\}$, $i \in I$, unabhängig sind (Satz 10.5). Es gilt weiter $\sigma(1_{A_i}) = \mathcal{A}_i$. Dies legt die folgende Definition nahe:

10.12 Definition Eine Familie $(X_i)_{i\in I}$ von Zufallsvariablen auf einem W-Raum $(\Omega, \mathcal{A}, P)$ mit Werten in $(\Omega_i, \mathcal{A}_i)$ heißt *unabhängig*, wenn die Familie

$$\big(\sigma(X_i)\big)_{i\in I} = \big(X_i^{-1}(\mathcal{A}_i)\big)_{i\in I}$$

von σ-Algebren unabhängig ist.

10.13 Satz *Für jedes $i = 1, \ldots, n$ sei*

$$X_i : (\Omega, \mathcal{A}, P) \to (\Omega_i, \mathcal{A}_i)$$

eine Zufallsvariable und $\mathcal{E}_i$ ein durchschnittstabiler Erzeuger von $\mathcal{A}_i$ mit $\Omega_i \in \mathcal{E}_i$. Die $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

für jede Auswahl $A_i \in \mathcal{E}_i$ gilt ($i = 1, \ldots, n$).

Beweis: $X_i^{-1}(\mathcal{E}_i)$ ist ein durchschnittstabiler Erzeuger von $\sigma(X_i)$, der $\Omega = X_i^{-1}(\Omega_i)$ enthält. Die Behauptung folgt dann aus Satz 10.5. $\square$

10.14 Korollar *Eine Familie von Zufallsgrößen $(X_i)_{i\in I}$ ist genau dann unabhängig, wenn für alle $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I$ und $t_1, \ldots, t_n \in \mathbb{R}$*

$$P(X_{i_1} \leq t_1, \ldots, X_{i_n} \leq t_n) = \prod_{j=1}^{n} P(X_{i_j} \leq t_j)$$

gilt.

Beweis: Dies folgt aus Definition 10.1, Satz 10.13 und der Tatsache, dass

$$\Big\{X_i^{-1}\big((-\infty,\, t]\big),\, t \in \mathbb{R}\Big\} \cup \Omega$$

ein durchschnittstabiles Erzeugendensystem von $X_i^{-1}(\mathcal{B})$ ist. □

10.15 Beispiel (siehe Beispiel 10.4 (b)) Die Folge $X_n := 1_{A_n}$ der Rademacher-Funktionen X_n ist nach $b(1, \frac{1}{2})$ verteilt. Also ist die Folge $(X_n)_n$ konvergent in Verteilung. Weiter gilt:

$$\begin{aligned} P(|X_m - X_n| \geq \delta) &= P(X_m = 1,\, X_n = 0) + P(X_m = 0,\, X_n = 1) \\ &= P(X_m = 1)P(X_n = 0) + P(X_m = 0)P(X_n = 1) \\ &= \frac{1}{2} \end{aligned}$$

für alle $n \neq m$ und δ mit $0 < \delta < 1$.

Mit Satz 9.12 und 9.13 folgt: Die Folge $(X_n)_n$ kann keine stochastisch konvergente Teilfolge enthalten!

10.16 Satz *Sei $(X_i)_{i \in I}$ eine unabhängige Familie $(\Omega_i,\, \mathcal{A}_i)$-wertiger Zufallsvariablen und*

$$f_i : (\Omega_i,\, \mathcal{A}_i) \to (\Omega_i',\, \mathcal{A}_i')$$

für jedes $i \in I$ eine messbare Abbildung. Dann ist auch die Familie $(f_i \circ X_i)_{i \in I}$ unabhängig.

Beweis: Für $A' \in \mathcal{A}_i'$ ist $(f_i \circ X_i)^{-1}(A') = X_i^{-1}\big(f_i^{-1}(A')\big)$ und somit $\sigma(f_i \circ X_i) \subset \sigma(X_i)$. Mit $\big(\sigma(X_i)\big)_{i \in I}$ ist daher auch $\big(\sigma(f_i \circ X_i)\big)_{i \in I}$ unabhängig. □

Die Unabhängigkeit von Zufallsvariablen ist eine wahrscheinlichkeitstheoretische Eigenschaft, also eine Eigenschaft ihrer Verteilungen:

10.17 Satz *Eine Familie von Zufallsvariablen $(X_i)_{i \in I}$ auf einem W-Raum $(\Omega,\, \mathcal{A},\, P)$ mit Werten in $(\Omega_i,\, \mathcal{A}_i)$ ist genau dann unabhängig, wenn ihre Verteilung die Produktverteilung ihrer Komponenten P^{X_i} ist:*

$$P^{(X_i)_{i \in I}} = \bigotimes_{i \in I} P^{X_i}.$$

($X_I = (X_i)_{i \in I}$ ist eine messbare Abbildung von $(\Omega,\, \mathcal{A})$ nach $(\prod \Omega_i,\, \bigotimes \mathcal{A}_i)$).

Beweis: Für jedes $J \subset I$ sei $X_J = (X_i)_{i \in J}$ und p_J die Projektion auf die Komponenten mit Index in J: $p_J \circ X_I = X_J$. Nach Satz 8.19 ist P^{X_I} genau dann das Produktmaß der P^{X_i}, $i \in I$, wenn für jedes $J = \{j_1, \ldots, j_n\} \in \mathcal{H}(I)$

$$P^{X_J} = p_J(P^{X_I}) = \bigotimes_{j \in J} P^{X_j}$$

gilt, also

$$P(X_{j_1} \in A_1, \ldots, X_{j_n} \in A_n) = \prod_{k=1}^{n} P^{X_{j_k}}(A_k) = \prod_{k=1}^{n} P(X_{j_k} \in A_k)$$

für messbare $A_1, \ldots, A_n$ gilt. Dies ist nach Satz 10.13 zur Unabhängigkeit der $(X_j)_{j \in J}$ und damit zur Unabhängigkeit der ganzen Familie $(X_i)_{i \in I}$ äquivalent. □

10.18 Korollar *Zu jeder Familie* $\big((\Omega_i, \mathcal{A}_i, P_i)\big)_{i \in I}$ *von W-Räumen existiert eine unabhängige Familie* $(X_i)_{i \in I}$ *von* $(\Omega_i, \mathcal{A}_i)$*-wertigen Zufallsvariablen auf einem geeigneten W-Raum* $(\Omega, \mathcal{A}, P)$*, so dass für jedes* $i \in I$ *gilt* $P_i = P^{X_i}$.

Beweis: Sei $(\Omega, \mathcal{A}, P) := \bigotimes_{i \in I}(\Omega_i, \mathcal{A}_i, P_i)$ und X_i die i-te Projektionsabbildung. Dann ist $(X_i)_{i \in I}$ die identische Abbildung auf Ω und hat $P = \bigotimes_{i \in I} P_i$ als Verteilung, was die Unabhängigkeit unter P beweist. □

10.19 Beispiel $\Omega_0 = \{0, 1\}$, $\mathcal{A}_0 = \mathcal{P}(\Omega_0)$, $A = \{1\}$, $P_0(A) = p$, $P_0(A^c) = q := 1 - p$, $0 \leq p \leq 1$.

Dann ist der Bernoulli-Versuch gegeben durch den W-Raum

$$\Omega = \Omega_0^{\mathbb{N}}, \quad \mathcal{A} = \mathcal{A}_0^{\mathbb{N}}, \quad P := P_0^{\mathbb{N}},$$

und besteht aus abzählbar oft unabhängigen Wiederholungen. Es sei

$$X_n(\omega) := \omega_n \quad \text{für } \omega = (\omega_n)_n \in \Omega\,.$$

Wir berechnen die Wahrscheinlichkeit dafür, dass unendlich oft zweimal hintereinander Kopf geworfen wird, bei einer fairen Münze $p = q = \frac{1}{2}$. A_n sei das Ereignis, dass beim n-ten und beim $(n+1)$-ten Wurf Kopf fällt. Dann ist $P(A_n) = \frac{1}{4}$ und

$$\sum_{n=1}^{\infty} P(A_{2n}) = +\infty\,.$$

$A := \limsup A_n$ interessiert uns. Es gilt $P(A) = 1$, denn $(A_{2n})_n$ ist eine Folge paarweise unabhängiger Ereignisse (sogar unabhängig) und

$$\limsup_{n \to \infty} A_{2n} \subset \limsup_{n \to \infty} A_n\,,$$

also wenden wir Satz 10.11 an.

Sei $(X_n)_n$ nun der BERNOULLI-Versuch wie in Beispiel 10.19. Dann ist $S_n := X_1 + \cdots + X_n$ $b(n,p)$-verteilt, denn mit der Unabhängigkeit ist $P((X_1, \ldots, X_n) = \omega) = p^k(1-p)^{n-k}$, wenn $S_n = k$ ist. Dann ist

$$\mathbb{E}(S_n) = n\mathbb{E}(X_1) = n\big(p \cdot 1 + (1-p) \cdot 0\big) = np$$

und dies ist eine deutlich schönere Herleitung als die bisher gegebene in Beispiel 7.28(a).

Satz 10.9 für von Zufallsvariablen erzeugte σ-Algebren besagt: Ist $(X_n)_n$ eine unabhängige Folge von Zufallsvariablen. Dann gilt für jedes terminale Ereignis

$$A \in \bigcap_{n=1}^{\infty} \sigma(X_m\,;\, m \geq n) =: \mathcal{T}_\infty$$

entweder $P(A) = 0$ oder $P(A) = 1$. Wir betrachten dazu ein Korollar:

10.20 Korollar *Es sei $(X_n)_{n\geq 1}$ eine unabhängige Folge reeller Zufallsvariablen. Dann ist jede $\mathcal{T}_\infty$-messbare numerische Zufallsvariable T fast sicher konstant, d.h. es existiert ein $\alpha \in \bar{\mathbb{R}}$ mit*

$$P(T = \alpha) = 1\,.$$

(T heißt manchmal terminale Funktion.)

Beweis: Sei $\gamma \in \bar{\mathbb{R}}$, dann ist $\{T \leq \gamma\} \in \mathcal{T}_\infty$ und somit $P(T \leq \gamma) = 0$ oder 1. Für $\gamma = +\infty$ ist $P(T \leq \gamma) = P(\Omega) = 1$.

Es sei α das Infimum in $\bar{\mathbb{R}}$ der somit nichtleeren Menge C aller $\gamma \in \bar{\mathbb{R}}$ mit $P(T \leq \gamma) = 1$. Dann gilt $\gamma_n \downarrow \alpha$ für eine geeignete antitone Folge $(\gamma_n)_n$ in C und mit $\{T \leq \gamma_n\} \downarrow \{T \leq \alpha\}$ ist $\alpha \in C$. α ist also das kleinste Element von C. Hieraus folgt $P(T < \alpha) = 0$ und $P(T = \alpha) = 1$. $\square$

Wir sammeln noch ein paar Rechenregeln:

10.21 Satz (Multiplikationssatz) *Seien $X_1, \ldots, X_n$ unabhängige reelle Zufallsvariablen. Dann gilt*

$$\mathbb{E}\Big(\prod_{i=1}^{n} X_i\Big) = \prod_{i=1}^{n} \mathbb{E}(X_i)$$

wenn alle $X_i \geq 0$ oder alle X_i integrierbar sind. Im zweiten Fall ist auch $\prod X_i$ integrierbar. Ist umgekehrt $\prod X_i$ integrierbar und verschwindet kein X_i fast sicher, so ist auch jedes X_i integrierbar.

Beweis: Nach Satz 10.17 ist $Q = \bigotimes_{i=1}^n P^{X_i}$ die gemeinsame Verteilung der $X_1, \ldots, X_n$. Es gilt

$$\begin{aligned}\mathbb{E}\Big(\big|\prod_{i=1}^n X_i\big|\Big) &= \int |x_1 \cdots x_n| \, Q(dx) \\ &= \int \cdots \int |x_1| \cdots |x_n| \, P^{X_1}(dx_1) \cdots P^{X_n}(dx_n) \\ &= \prod_{i=1}^n \int |x_i| \, P^{X_i}(dx_i) = \prod_{i=1}^n \mathbb{E}(|X_i|)\end{aligned}$$

nach Satz 5.25 und Satz 8.12 (Fubini).

Also folgt die Behauptung für $X_i \geq 0$ und die Integrierbarkeit von $\prod X_i$ für den Fall, dass alle X_i integrierbar sind. Nach Fubini bleibt dann die obige Rechnung richtig, wenn die Absolut-Striche fehlen.

Gilt $\mathbb{E}\Big(\big|\prod_{i=1}^n X_i\big|\Big) < \infty$ und $\mathbb{E}(|X_i|) > 0$, $i = 1, \ldots, n$, so ist

$$\prod_{i=1}^n \mathbb{E}(|X_i|) = \mathbb{E}\Big(\big|\prod_{i=1}^n X_i\big|\Big) < \infty$$

und kein Faktor Null, also $\mathbb{E}(|X_i|) < \infty$, d.h. jedes X_i ist integrierbar. □

10.22 Korollar *Zwei unabhängige Zufallsgrößen X und Y mit endlichem Erwartungswert sind unkorreliert.*

Beweis: Da X und Y unabhängig sind, sind auch $X - \mathbb{E}(X)$ und $Y - \mathbb{E}(Y)$ unabhängig (Satz 10.16) und ihr Erwartungswert ist Null. Somit folgt mit 10.21 die Behauptung. □

Im Fall des Bernoulli-Versuchs sind die $(X_n)_n$ also paarweise unkorreliert, somit gilt $\mathrm{Var}(S_n) = n \,\mathrm{Var}(X_1)$ (siehe Definition 7.22 und Folgerung). Nun gilt $\mathrm{Var}(X_1) = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2$ nach Satz 7.16 und hier ist $\mathbb{E}X_1^2 = \mathbb{E}X_1 = p$, also $\mathrm{Var}(X_1) = p - p^2 = p(1-p)$, also $\mathrm{Var}(S_n) = np(1-p)$, wieder eine deutlich schönere Herleitung als die bisher gegebene in Beispiel 7.28(a).

Wir untersuchen die Verteilung der Summe zweier unabhängiger Zufallsgrößen X, Y.

10.23 Definition Die *Faltung* zweier W-Maße P_1 und P_2 auf $(\mathbb{R}, \mathcal{B})$ ist das Bildmaß

$$P_1 * P_2 := (P_1 \otimes P_2) \circ S^{-1},$$

wobei $S : \mathbb{R}^2 \to \mathbb{R}$ definiert ist als $S(x_1, x_2) = x_1 + x_2$ (kann analog für allgemeine Bildräume definiert werden).

X und Y seien Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$, die unabhängig sind. Seien $\mu = P^X$ und $\nu = P^Y$, dann ist

$$P(X+Y<t) = \int_{\mathbb{R}^2} f_t(x,\, y)\, (\mu \otimes \nu)\, d(x,\, y)$$

mit

$$f_t(x,\, y) = 1_{\{x+y<t\}} = 1_{(-\infty,\, t-y)}(x)\,,$$

da $\mu \otimes \nu$ die Verteilung von $(X,\, Y)$ ist. Nach Fubini ist

$$\begin{aligned} P(X+Y<t) &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} 1_{(-\infty,\, t-y)}(x)\, \mu(dx) \Big) \nu(dy) \\ &= \int_{\mathbb{R}} F_X(t-y)\, \nu(dy) \end{aligned}$$

wobei F_X die Verteilungsfunktion von X ist.

10.24 Satz *X und Y seien unabhängige Zufallsgrößen mit* Lebesgue-*Dichten f und g. Dann hat $X+Y$ die* Lebesgue-*Dichte*

$$h_{X+Y}(z) = \int f(y)\, g(z-y)\, \lambda(dy) = \int f(z-y)\, g(y)\, \lambda(dy)\,,$$

XY die Lebesgue-*Dichte*

$$h_{XY}(z) = \int \frac{1}{|y|} f\Big(\frac{z}{y}\Big)\, g(y)\, \lambda(dy) = \int \frac{1}{|y|} f(y)\, g\Big(\frac{z}{y}\Big)\, \lambda(dy)$$

und X/Y die Lebesgue-*Dichte*

$$h_{X/Y}(z) = \int |y|\, f(zy)\, g(y)\, \lambda(dy) = \frac{1}{z^2} \int |y|\, f(y)\, g\Big(\frac{y}{z}\Big)\, \lambda(dy)\,,$$

wobei $P(Y \neq 0) = 1$ durch die λ-Stetigkeit von Y garantiert wird.

Beweis: Wir beweisen nur die Formel für h_{X+Y}, der Rest ist eine Übung. Wir verwenden die Translationsinvarianz von λ.

$$\begin{aligned} P(X+Y<t) &= \int_{\mathbb{R}} \Big(\int_{(-\infty,\, t-y)} f(x)\, \lambda(dx) \Big) g(y)\, \lambda(dy) \\ &= \int_{\mathbb{R}} \Big(\int_{(-\infty,\, t)} f(x-y)\, \lambda(dx) \Big) g(y)\, \lambda(dy) \\ &= \int_{(-\infty,\, t)} \Big(\int_{\mathbb{R}} f(x-y)\, g(y)\, \lambda(dy) \Big) \lambda(dx)\,. \end{aligned}$$

□

KAPITEL 11

Starkes Gesetz der großen Zahlen

Im Bernoulli-Experiment aus Beispiel 10.19 nimmt $S_n := \sum_{i=1}^n X_i$ nur die Werte $0, 1, \ldots, n$ an und gibt die Zahl der Erfolge bei den ersten n Ausführungen an. Die relative Häufigkeit $n^{-1}S_n$ sollte mit großer Wahrscheinlichkeit gegen p streben. Dieses vage Gefühl soll nun präzisiert werden.

Für $\omega = (0, 0, \ldots)$ bzw. $\omega = (1, 1, \ldots)$ ist $n^{-1}S_n(\omega) = 0$ bzw. $n^{-1}S_n(\omega) = 1$ für alle $n \in \mathbb{N}$, also konvergiert $n^{-1}S_n(\omega)$ für $n \to \infty$ offenbar nicht für jedes $\omega \in \Omega$. Konvergiert diese Größe stochastisch oder gar fast sicher? Im Bernoulli-Experiment hat X_n die Verteilung $b(1, p)$ und den Erwartungswert $\mathbb{E}(X_n) = p$. Wir fragen also:

Gilt

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \big(X_i - \mathbb{E}(X_i)\big) = 0 \tag{11.1}$$

im Sinne der stochastischen bzw. der fast sicheren Konvergenz bzgl. P? Man sagt, dass eine Folge $(X_n)_n$ integrierbarer reeller Zufallsvariablen dem *schwachen* bzw. dem *starken Gesetz der großen Zahlen* genügt, wenn (11.1) im Sinne der stochastischen bzw. der P-fast sicheren Konvergenz gilt. Bei einer beliebigen Folge identisch verteilter Zufallsgrößen $(X_n)_n$ ist $\mathbb{E}(X_1) = \mathbb{E}(X_n)$ für alle $n \in \mathbb{N}$, wenn $\mathbb{E}(X_1)$ existiert, und (11.1) wird zu

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}(X_1) \ . \tag{11.2}$$

Man sagt daher allgemeiner, dass eine Folge reeller Zufallsvariablen $(X_n)_n$

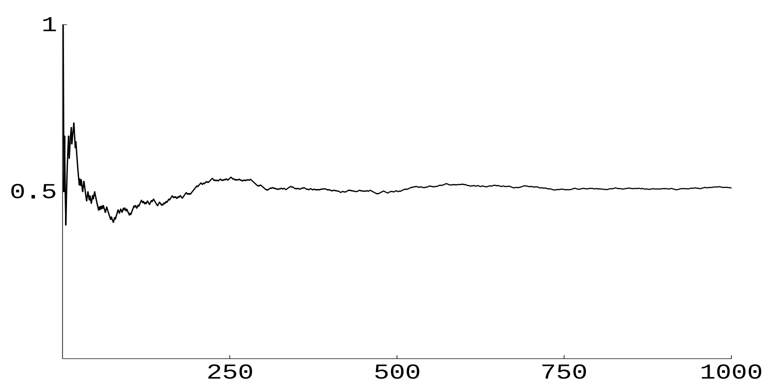


Abb. 11.1: Eine Simulation von $n^{-1}S_n$ bei der fairen Münze.

dem schwachen bzw. dem starken Gesetz der großen Zahlen genügt, wenn (11.2) gilt für $\mathbb{E}(X_1)$ ersetzt durch eine reelle Zahl $\mu \in \mathbb{R}$ (wieder im Sinne der stochastischen bzw. der P-fast sicheren Konvergenz). Natürlich folgt aus der Gültigkeit eines starken Gesetzes die des korrespondierenden schwachen Gesetzes, siehe Satz 9.9.

Dem schwachen Gesetz hatten wir uns in Satz 7.30 bereits gewidmet. Der dort gegebene Beweis führt unmittelbar zu

11.1 Satz (von Khintchine) *Gilt für eine Folge $(X_n)_n$ integrierbarer und paarweise unkorrelierter reeller Zufallsvariablen*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) = 0 \ ,$$

so genügt die Folge dem schwachen Gesetz der großen Zahlen.

Beweis: Aus der Voraussetzung folgt, dass alle X_i quadratisch integrierbar sind. Es gilt

$$\mathrm{Var}\Big(\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i),$$

also

$$\mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))\Big) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i),\ \ n \in \mathbb{N}.$$

Die Behauptung folgt mit der TSCHEBYSCHEV-Ungleichung (Satz 7.17(b)). □

Kommt man auch ohne quadratische Integrierbarkeit aus? Wir notieren hier das folgende Resultat ohne Beweis:

11.2 Satz *Sind die $(X_n)_n$ unabhängig und identisch verteilt, so genügt $(X_n)_n$ (ohne die Annahme der Integrierbarkeit!) genau dann dem schwachen Gesetz der großen Zahlen mit Limes* 0, *wenn*

$$\lim_{n\to\infty} n\, P\big(|X_1| > n\big) = 0$$

und

$$\lim_{n\to\infty} \int_{\{|X_1|\geq n\}} X_1 \, dP = 0$$

gilt.

Für einen Beweis siehe zum Beispiel das Buch von GÄNSSLER und STUTE, Wahrscheinlichkeitstheorie, Satz 2.1.11.

Das 0-1-Gesetz, Satz 10.9, lässt die Frage nach der Gültigkeit des starken Gesetzes der großen Zahlen in der folgenden Sicht erscheinen: Es sei $(X_n)_n$ eine Folge unabhängiger Zufallsgrößen, $\mathcal{T}_\infty$ definiert wie in 10.20 und $S_n := \sum_{j=1}^n X_j$. Dann gilt:

11.3 Lemma *Sei $(\tau_n)_n$ eine Nullfolge reeller Zahlen, dann sind $\liminf_{n\to\infty} \tau_n S_n$ und $\limsup_{n\to\infty} \tau_n S_n$ $\mathcal{T}_\infty$-messbare Zufallsgrößen, also fast sicher konstant (Konstante in $[-\infty, +\infty]$).*

Beweis: Für $m \in \mathbb{N}$ gilt

$$\limsup_{n\to\infty} \tau_n S_n = \limsup_{n\to\infty} \tau_n \left(\sum_{j=1}^m X_j + \sum_{j=m+1}^n X_j \right) = \limsup_{n\to\infty} \tau_n \left(\sum_{j=m+1}^n X_j \right) .$$

Die Zufallsgröße auf der rechten Seite ist $\sigma(X_n, n \geq m+1)$-messbar für jedes $m \in \mathbb{N}$, also $\mathcal{T}_\infty$-messbar. Für $\liminf_{n\to\infty}$ folgt die Aussage analog. Mit Korollar 10.20 folgt die Behauptung. $\square$

Es sei nun $A = \{\omega \in \Omega : \lim_{n\to\infty} \tau_n S_n(\omega) = 0\}$. Dann ist

$$A = \{\limsup_{n\to\infty} \tau_n S_n = 0\} \cap \{\liminf_{n\to\infty} \tau_n S_n = 0\}$$

ein Ereignis und nach Lemma 11.3 ist $A \in \mathcal{T}_\infty$. Also ist $P(A)$ gleich 0 oder 1. Ist jedes X_n integrierbar, so ist $(X_n - \mathbb{E}(X_n))_n$ auch eine unabhängige Folge und für $\tau_n = \frac{1}{n}$ gilt somit, dass

$$P\Big(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}(X_i)) = 0 \Big)$$

entweder 0 oder 1 ist. Das starke Gesetz der großen Zahlen gilt dann, wenn diese Wahrscheinlichkeit 1 ist. Tatsächlich kann das starke Gesetz der großen Zahlen für eine große Klasse von Folgen von Zufallsgrößen bewiesen werden:

11.4 Satz (von Etemadi, 1981) *Jede Folge $(X_n)_n$ reeller, identisch verteilter und paarweise unabhängiger Zufallsvariablen genügt genau dann dem starken Gesetz der großen Zahlen mit Limes μ, wenn X_1 integrierbar ist mit $\mu = \mathbb{E}(X_1)$.*

Kolmogorov publizierte 1930 das starke Gesetz für unabhängige Zufallsvariablen:

11.5 Korollar (von Kolmogorov, 1930) *Jede unabhängige Folge identisch verteilter, integrierbarer reeller Zufallsvariablen genügt dem starken Gesetz der großen Zahlen.*

Beweis des Satzes von ETEMADI*:* „$\Rightarrow$“: Sei $S_n := \sum_{j=1}^n X_j$ und $\frac{S_n}{n} \to \mu$ f.s. für ein $\mu \in \mathbb{R}$. Es folgt

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \to 0 \text{ f.s.}$$

und somit

$$P\big(|X_n| > n \text{ unendlich oft}\big) = P\big(\limsup_{n\to\infty}\{|X_n| > n\}\big) = 0 \ .$$

Also konvergiert $\sum_{n\geq 1} P(|X_n| > n)$ nach BOREL-CANTELLI (Satz 10.11), denn die X_n sind paarweise unabhängig. Da die X_n auch identisch verteilt sind, folgt nach Beispiel 8.18

$$\mathbb{E}|X_1| \leq 1 + \sum_{n\geq 1} P(|X_1| > n) = 1 + \sum_{n\geq 1} P(|X_n| > n) < \infty \ .$$

„$\Leftarrow$“: Natürlich ist dies das Herzstück des Satzes. Es wird behauptet: $\lim_{n\to\infty} \frac{1}{n} S_n = \mu = \mathbb{E}(X_1)$ P-fast sicher. Mit $(X_n)_n$ genügen auch $(X_n^+)_n$ und $(X_n^-)_n$ den Voraussetzungen des Satzes. Es kann daher ohne Einschränkung $X_n \geq 0$ für alle $n \in \mathbb{N}$ angenommen werden. Wir betrachten die gestutzte Folge

$$Y_n := X_n 1_{\{X_n \leq n\}} \ , \ n \in \mathbb{N} \ .$$

Es geht die ursprünglich gemeinsame Verteilung verloren. Wir gewinnen aber die quadratische Integrierbarkeit, denn

$$\mathbb{E}(Y_n^2) = \int_{[0,n]} x^2 \, P^{X_1}(x) < \infty \ .$$

Es genügt zu zeigen, dass $\frac{1}{n}\sum_{i=1}^n Y_i$ fast sicher gegen μ konvergiert. Dazu betrachte

$$\begin{aligned} \sum_{n\geq 1} P(X_n \neq Y_n) &= \sum_{n\geq 1} P(X_n > n) = \sum_{n\geq 1} P(X_1 > n) \\ &\leq \mathbb{E}X_1 < \infty, \end{aligned}$$

wobei wir wieder Beispiel 8.18 verwendet haben. Also folgt nach BOREL-CANTELLI $P(X_n \neq Y_n \text{ unendlich oft}) = 0$. Somit hat das Ereignis A aller $\omega \in \Omega$ mit $X_n(\omega) \neq Y_n(\omega)$ für höchstens endlich viele $n \in \mathbb{N}$ Wahrscheinlichkeit 1. Also kann aus der fast sicheren Konvergenz von $\frac{1}{n}\sum Y_i$ gegen μ auf die fast sichere Konvergenz von $\frac{1}{n}\sum X_i$ gegen μ geschlossen werden.

Es sei nun $T_n := \frac{1}{n}\sum_{i=1}^n Y_i$, $n \in \mathbb{N}$. Zu $\alpha > 1$ setzen wir $k_n := [\alpha^n] := \sup\{m \in \mathbb{N}_0 : \ m \leq \alpha^n\}$. Zu $\varepsilon > 0$ liefert die TSCHEBYSCHEV-Ungleichung

$$\sum_{n\geq 1} P(|T_{k_n} - \mathbb{E}(T_{k_n})| > \varepsilon) \leq \frac{1}{\varepsilon^2}\sum_{n\geq 1}\frac{1}{k_n^2}\sum_{i=1}^{k_n} \mathrm{Var}(Y_i) \ ,$$

denn die $(Y_i)_i$ sind paarweise unabhängig, also paarweise unkorreliert. Wir nutzen $\mathrm{Var}(Y_i) \leq \mathbb{E}(Y_i^2)$ und vertauschen mit Fubini die Summationsreihenfolge (alle Summanden sind nicht negativ) und erhalten:

$$\sum_{n\geq 1} P(|T_{k_n} - \mathbb{E}(T_{k_n})| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \mathbb{E}(Y_i^2) \sum_{n: k_n \geq i} \frac{1}{k_n^2}.$$

Es sei n_i die kleinste natürliche Zahl n mit $k_n \geq i$. Wir schätzen $\sum_{n=n_i}^{\infty} \frac{1}{k_n^2}$ ab: Aus $[\alpha^n] \geq \alpha^n/2$ für alle $n \geq 1$ folgt

$$\sum_{n=n_i}^{\infty} \frac{1}{k_n^2} \leq 4 \sum_{n=n_i}^{\infty} \alpha^{-2n} = 4\alpha^{-2n_i} \sum_{n=n_i}^{\infty} \alpha^{-2(n-n_i)} \leq \frac{4}{i^2} \frac{1}{(1-\alpha^{-2})} ,$$

und somit

$$\sum_{n\geq 1} P(|T_{k_n} - \mathbb{E}(T_{k_n})| > \varepsilon) \leq \frac{4}{\varepsilon^2(1-\alpha^{-2})} \sum_{i=1}^{\infty} \frac{\mathbb{E}(Y_i^2)}{i^2} .$$

Nun wollen wir die Reihe auf der rechten Seite untersuchen. Dazu verwenden wir Beispiel 8.18 wie folgt. Für $a > 0$ und jede nicht negative Zufallsgröße X ist

$$\mathbb{E}\big(X^2 1_{\{X \leq a\}}\big) = \int_0^a 2tP(X > t, X \leq a)\, dt \leq \int_0^a 2tP(X > t)\, dt.$$

Also erhalten wir

$$\begin{aligned}
\sum_{i=1}^{\infty} \frac{\mathbb{E}(Y_i^2)}{i^2} &\leq \sum_{i=1}^{\infty} \frac{1}{i^2} \int_0^i 2tP(X_1 > t)\, dt \\
&= \sum_{i=1}^{\infty} \sum_{k=1}^{i} \frac{1}{i^2} \int_{k-1}^{k} 2tP(X_1 > t)\, dt \\
&= \sum_{k\geq 1} \Big(\int_{k-1}^{k} 2tP(X_1 > t)\, dt \sum_{i\geq k} \frac{1}{i^2} \Big) \\
&\leq 2 + \sum_{k\geq 2} \frac{2k-1}{k-1} P(X_1 > k-1) \leq 2 + 3\mathbb{E}(X_1) < \infty.
\end{aligned}$$

Wir müssen nur noch die letzte Abschätzung begründen. Wir verwenden

$$\sum_{i\geq 1} \frac{1}{i^2} \leq 1 + \sum_{i\geq 1} \frac{1}{i(i+1)} = 2$$

und

$$\sum_{i\geq k} \frac{1}{i^2} \leq \sum_{i\geq k-1} \frac{1}{i(i+1)} = \frac{1}{k-1}$$

und für $k \geq 2$

$$\int_{k-1}^{k} 2tP(X_1 > t)\,dt \leq P(X_1 > k-1) \int_{k-1}^{k} 2t\,dt = (2k-1)P(X_1 > k-1).$$

Also konvergiert $T_{k_n} - \mathbb{E}(T_{k_n})$ f.s. gegen Null gemäß Satz 9.9.

Nach dem Satz von der monotonen Konvergenz folgt $\mathbb{E}(Y_i) = \mathbb{E}\big(X_1\, 1_{\{X_1 \leq i\}}\big) \to \mu = \mathbb{E}(X_1)$, und somit $\mathbb{E}(T_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y_i) \to \mu$ (siehe Analysis). Also impliziert $T_{k_n} - \mathbb{E}(T_{k_n}) \to 0$ f.s. $T_{k_n} \to \mu$ fast sicher.

Zu untersuchen bleiben die Werte in $\mathbb{N} \setminus \{k_n, n \geq 1\}$. k_n geht isoton gegen $+\infty$, also gibt es zu jedem $m \in \mathbb{N}$ mit $m > k_1$ genau ein $n \in \mathbb{N}$ mit $k_n < m \leq k_{n+1}$. Da alle $X_i \geq 0$, folgt

$$\frac{k_n}{k_{n+1}} T_{k_n} \leq T_m \leq T_{k_{n+1}} \frac{k_{n+1}}{k_n}\ .$$

Beachte

$$\frac{1}{\alpha} - \frac{1}{\alpha^{n+1}} = \frac{\alpha^n - 1}{\alpha^{n+1}} \leq \frac{[\alpha^n]}{[\alpha^{n+1}]} \leq \frac{\alpha^n}{\alpha^{n+1}-1} = \frac{1}{\alpha - \frac{1}{\alpha^n}},$$

also $\lim_{n\to\infty} \frac{k_{n+1}}{k_n} = \alpha$ für $\alpha > 1$. Somit folgt

$$\frac{\mu}{\alpha} \leq \liminf_{m\to\infty} T_m \leq \limsup_{m\to\infty} T_m \leq \alpha\mu$$

fast sicher. Da $\alpha > 1$ beliebig vorgegeben war, folgt $T_m \to \mu$ fast sicher, was zu zeigen war. □

11.6 Korollar *Es sei $(X_n)_n$ eine Folge identisch verteilter und paarweise unabhängiger Zufallsgrößen. Aus $\mathbb{E}(X_1) = \infty$ bzw. $-\infty$ folgt $n^{-1}S_n \to +\infty$ bzw. $-\infty$ fast sicher.*

Beweis: Wir zeigen $\mathbb{E}(X_1) = \infty \Rightarrow n^{-1}S_n \to \infty$ fast sicher:

Es sei $X_n^c := X_n 1_{\{X_n \leq c\}}$, $c > 0$. $(X_n^c)_n$ ist dann paarweise unabhängig, identisch verteilt und integrierbar. Es gilt weiter $S_n \geq S_n^c := \sum_{j=1}^{n} X_j^c$, $n \in \mathbb{N}$, $c > 0$, also mit Satz 11.4

$$\liminf_{n\to\infty} \frac{S_n}{n} \geq \lim_{n\to\infty} \frac{S_n^c}{n} = \mathbb{E}(X_1^c) \quad \text{f.s.}$$

für alle $c > 0$, und somit folgt mit $\mathbb{E}(X_1^c) \nearrow \mathbb{E}(X_1) = \infty$ das Gewünschte. □

Das starke Gesetz kann auch für nicht notwendig identisch verteilte Zufallsvariablen hergeleitet werden. Der sogenannte „klassische"Weg zum starken Gesetz (Kolmogorovsches Kriterium, 1928) führt über eine stochastische Ungleichung:

11.7 Satz (Kolmogorov-Ungleichung) *Es sei* $(X_n)_n$ *eine Folge unabhängiger Zufallsgrößen mit* $\mathbb{E}(X_i) = 0$ *für alle* $i \in \mathbb{N}$. *Dann gilt für jedes* $\varepsilon > 0$

$$P\big(\max_{1 \le k \le n} |S_k| \ge \varepsilon\big) \le \frac{1}{\varepsilon^2} \sum_{k=1}^{n} \mathrm{Var}(X_k) \ .$$

Beweis: Es sei $A_1 := \{|S_1| \ge \varepsilon\}$ und

$$A_{k+1} := \{|S_{k+1}| \ge \varepsilon \text{ und } \max_{1 \le l \le k} |S_l| < \varepsilon\}.$$

Dann sind die A_k's disjunkt und $B_n := \{\max_{1 \le k \le n} |S_k| \ge \varepsilon\} = \cup_{k=1}^n A_k$. Nun gilt für $1 \le k < n$

$$S_n^2 - S_k^2 = (S_n - S_k)^2 + 2(S_n - S_k)S_k \ge 2(S_n - S_k)S_k \ .$$

Da $\mathbb{E}(S_n - S_k) = 0$ und $S_n - S_k$ unabhängig von S_k, folgt $\mathbb{E}(S_n^2 - S_k^2) \ge 0$, also $\mathbb{E}(S_n^2) \ge \mathbb{E}(S_k^2)$. Diese Ungleichung verwenden wir nun wie folgt:

$$\begin{aligned}
\mathbb{E}(S_n^2 1_{B_n}) &= \sum_{k=1}^{n} \mathbb{E}(S_n^2 1_{A_k}) \\
&\ge \sum_{k=1}^{n} \mathbb{E}(S_k^2 1_{A_k}) \\
&\ge \varepsilon^2 \sum_{k=1}^{n} P(A_k) = \varepsilon^2 P(B_n) \ .
\end{aligned}$$

Beachte nun noch $\mathbb{E}(S_n^2) = \sum_{k=1}^n \mathbb{E}(X_k^2) = \sum_{k=1}^n \mathrm{Var}(X_k)$. Damit ist die Ungleichung bewiesen. □

11.8 Satz *Genügt eine unabhängige Folge* $(X_n)_n$ *von Zufallsgrößen mit* $\mathbb{E}(X_i) = 0$ *für alle* $i \in \mathbb{N}$ *der Bedingung* $\sum_{n \ge 1} \mathrm{Var}(X_n) < \infty$, *so ist die Reihe* $\sum_{n \ge 1} X_n$ *f.s. endlich.*

Beweis: Wir betrachten

$$\begin{aligned}
\lim_{m \to \infty} P\Big(\sup_{n \ge m} |S_n - S_m| > \varepsilon\Big) &= \lim_{m \to \infty} \lim_{M \to \infty} P\Big(\max_{m \le n \le M} |S_n - S_m| > \varepsilon\Big) \\
&\le \lim_{m \to \infty} \lim_{M \to \infty} \frac{1}{\varepsilon^2} \sum_{k=m}^{M} \mathrm{Var}\, X_k = 0.
\end{aligned}$$

Hierbei verwenden wir die Voraussetzung für die Varianzen sowie die Ungleichung von Kolmogorov, Satz 11.7. Somit folgt die Behauptung aus Satz 9.4. □

Das anschließende Lemma wird die Verbindung zwischen der Konvergenz zufälliger Reihen und dem starken Gesetz der großen Zahlen herstellen:

11.9 Korollar (von Kronecker) *Sei $(c_n)_n$ eine Zahlenfolge und $(a_n)_n$ eine aufsteigende Folge mit $a_n > 0$ für alle $n \in \mathbb{N}$. Dann gilt:*

$$\sum_{n \geq 1} \frac{c_n}{a_n} < \infty \Rightarrow \lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} c_k = 0 \ .$$

Beweis: Wird in den Übungen besprochen.

11.10 Satz (Kolmogorovsches Kriterium) *Es sei $(X_n)_n$ eine unabhängige Folge von Zufallsgrößen und $(a_n)_n$ eine Zahlenfolge mit $0 < a_n \nearrow \infty$. Dann folgt aus $\sum_{n\geq 1} a_n^{-2} \operatorname{Var}(X_n) < \infty$:*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} (X_k - \mathbb{E}(X_k)) = 0 \quad \textit{f.s.}$$

Beweis: Die Folge $(Y_n)_n$ mit $Y_n := a_n^{-1}(X_n - \mathbb{E}(X_n))$, $n \geq 1$, genügt den Voraussetzungen in Satz 11.8. Somit ist $\sum_{n\geq 1} Y_n$ fast sicher endlich. Daraus folgt die Behauptung mit Hilfe des Lemmas von Kronecker, 11.9. $\square$

Für $a_n = n$ ergibt sich die Aussage von Korollar 11.5, allerdings nur unter der stärkeren Voraussetzung $X_n \in \mathcal{L}_2$. Satz 11.10 kann im Fall identisch verteilter Zufallsgrößen auch für $a_n = n^{\delta+1/2}$ oder $a_n = n^{1/2}(\log n)^{\delta+1/2}$ für $\delta > 0$, nicht jedoch für $a_n = n^{1/2}(\log n)^{1/2}$ angewandt werden. Bei der Wahl $\delta = 1/2$ erhalten wir zum Beispiel

$$\lim_{n\to\infty} \frac{S_n}{\sqrt{n}\log n} = 0 \quad P\text{-f.s.}$$

(im Fall $\mathbb{E}(X_1) = 0$).

Wir fassen diese Beobachtungen zusammen: Es sei $S_n := \sum_{j=1}^n X_j$, wobei $(X_n)_n$ eine Folge identisch verteilter, reeller, quadratisch integrierbarer Zufallsgrößen mit $\mathbb{E}(X_1) = 0$ ist. Dann ist $\lim_{n\to\infty} \frac{S_n}{n} = 0$ und $\lim_{n\to\infty} \frac{S_n}{\sqrt{n}\log n} = 0$ P-f.s. Für jede isotone Folge $(a_n)_n \nearrow \infty$ wissen wir nach Lemma 11.3, dass es ein $\tau \in \bar{\mathbb{R}}$ gibt mit

$$P\Big(\limsup_{n\to\infty} \frac{1}{a_n} S_n = \tau \Big) = 1 \ .$$

Ebenfalls ist $P\Big(\liminf_{n\to\infty} \frac{1}{a_n} S_n = -\tau \Big) = 1$ aus Symmetriegründen. Für die schnell wachsenden Folgen $a_n = n$ und $a_n = \sqrt{n}\log n$ folgt $\tau = 0$. Für $a_n = \sqrt{n}$

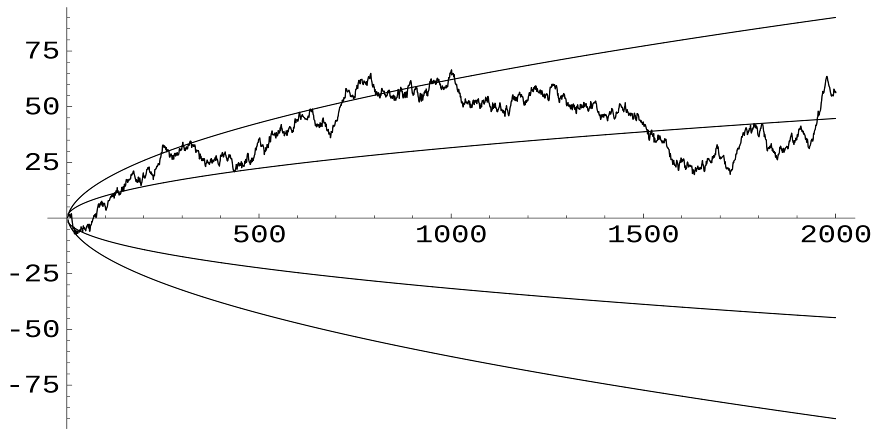


Abb. 11.2: Realisierung $\big((S_n(\omega)\big)_{1\le n\le 2000}$ bei unabhängigen, $N(0,1)$-verteilten Zuwächsen zusammen mit den Abbildungen $n \mapsto \pm\sqrt{n}$ und $n \mapsto \pm\sqrt{2n\log\log n}$.

werden wir $\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = +\infty$ P-f.s. sehen (als Folge aus dem zentralen Grenzwertsatz).

Kann man $(a_n)_n$ so bestimmen, dass τ reell und > 0 ist? Dann tritt das Ereignis $\{S_n \ge \eta a_n\}$ für $\eta < \tau$ unendlich oft und für $\eta > \tau$ nur endlich oft ein, jeweils mit Wahrscheinlichkeit 1! Tatsächlich gilt:

11.11 Satz (Gesetz vom iterierten Logarithmus) *Es sei $(X_n)_n$ eine Folge unabhängiger, identisch verteilter, quadratisch integrierbarer Zufallsgrößen mit $\mathbb{E}(X_1) = 0$.*

Dann gelten

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = \sigma \text{ fast sicher}$$

und

$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = -\sigma \text{ fast sicher}$$

mit $\sigma^2 := \mathrm{Var}(X_n)$.

Wir führen den Beweis hier nicht.

Das folgende numerische Gedankenexperiment gibt Auskunft über $\log(\log n)$: Jemand werfe in jeder Sekunde eine Münze einmal, n bezeichne die Anzahl der Würfe. Ist dann x der Zeitpunkt, an dem das Experiment begonnen hat, welches er nun abbricht, so gilt für $\log(\log n)$:

x	$\log(\log n)$
vor einer Stunde	$2,103$
beim Tode CAESARS	$3,214$
Geburt des Universums	$3,706$

Wir betrachten abschließend eine Anwendung des starken Gesetzes der großen Zahlen. Wie in Beispiel 10.4(b) bzw. 10.15 sei $(X_n)_n$ die Folge der unabhängigen RADEMACHER-Funktionen $X_n = 1_{A_n}$ mit

$$A_n = \bigcup_{k=0}^{2^{n-1}-1} \Big[\frac{2k}{2^n}, \frac{2k+1}{2^n}\Big) .$$

X_n ist $b(1,1/2)$ verteilt und nach dem starken Gesetz ist $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = 1/2$ P-f.s.

Jedes $\omega \in [0,1)$ besitzt bezüglich $g \geq 2$, $g \in \mathbb{N}$, genau eine g-adische Entwicklung

$$\omega = \sum_{n=1}^{\infty} \xi_n g^{-n} \ , \ \xi_n(\omega) \in \{0,\dots,g-1\} \tag{11.3}$$

(Nicht schließlich alle ξ_n sind gleich $g-1$).

$S_n^{\varepsilon,g}(\omega)$ sei die Anzahl aller $i = 1,\dots,n$ mit $\xi_i(\omega) = \varepsilon$ in der g-adischen Entwicklung von ω, $\varepsilon \in \{0,\dots,g-1\}$.

ω heißt *g-normal*, wenn

$$\lim_{n\to\infty} \frac{1}{n} S_n^{\varepsilon,g}(\omega) = \frac{1}{g}$$

für $\varepsilon = 0,\dots,g-1$ gilt.

ω heißt *absolut normal*, wenn sie für *alle* $g = 2,3,\dots$ g-normal ist.

Wir betrachten den Fall $g = 2$. Es ist dann

$$S_n^{0,2}(\omega) = \sum_{i=1}^n X_i, \quad S_n^{1,2} = n - S_n^{0,2}, \ \ n \in \mathbb{N}.$$

Also sind P-fast alle Zahlen $\omega \in \Omega$ 2-normal. Tatsächlich sind P-fast alle $\omega \in \Omega$ auch g-normal für $g > 2$. Betrachte dazu

$$\xi_n : \Omega \to \{0,1,\dots,g-1\},$$

definiert durch (11.3). Dann gilt

$$\{\xi_n = \varepsilon\} = \bigcup_{k=0}^{g^{n-1}-1} \Big[\frac{kg+\varepsilon}{g^n}, \frac{kg+\varepsilon+1}{g^n}\Big)$$

für $\varepsilon \in \{0, \ldots, g-1\}$. Also sind alle ξ_n Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ mit $P(\xi_n = \varepsilon) = \frac{1}{g}$. Analog zu den Überlegungen in Beispiel 10.4(b) folgt, dass $(\xi_n)_n$ eine unabhängige Folge ist. Nun definieren wir für jedes $\varepsilon \in \{0, \ldots, g-1\}$

$$X_n^\varepsilon := 1_{\{\xi_n = \varepsilon\}}.$$

Dann sind diese alle $b(1, \frac{1}{g})$-verteilt und es gilt

$$S_n^{\varepsilon,g} = \sum_{i=1}^n X_i^\varepsilon.$$

Das starke Gesetz der großen Zahlen liefert nun

$$\lim_{n\to\infty} \frac{1}{n} S_n^{\varepsilon,g} = \frac{1}{g} \quad P-\text{f.s.}.$$

Wir haben also den folgenden Satz bewiesen:

11.12 Satz *In Bezug auf das* Lebesgue*-Maß auf* $[0,1)$ *sind fast alle Zahlen* $\omega \in [0,1)$ *absolut normal.*

11.13 Bemerkung Champernowne hat 1933 gezeigt, dass

$$0,12345678910111213141516171819 2021\ldots$$

10-normal ist. Die Frage, ob zum Beispiel $\sqrt{2}$, $\log 2$, e oder π normal sind, ist unbeantwortet. Man kennt kein konkretes Beispiel einer absolut normalen Zahl. Eine g-normale Zahl ist im Allgemeinen nicht absolut normal (dies haben Cassels 1959 und Schmidt 1962 beobachtet). Eine g-normale Zahl ist stets g^p-normal für jedes $p \in \mathbb{N}$ (bewiesen von Hlawka, 1979).

KAPITEL 12

Der zentrale Grenzwertsatz

Für jedes $n \in \mathbb{N}$ seien $X_{n1}, \dots, X_{nk_n}$ unabhängige Zufallsgrößen, definiert auf einem W-Raum $(\Omega_n, \mathcal{A}_n, P_n)$, mit endlichen Varianzen $\sigma_{nk}^2 := \mathrm{Var}(X_{nk})$, $k = 1, \dots, k_n$. Betrachte

$$S_n := X_{n1} + \cdots + X_{nk_n} \; .$$

Bei der Familie $(X_{nk})_{n \geq 1}^{k=1,\dots,k_n}$ spricht man von einem *Dreiecksschema* mit gegen unendlich strebender Zeilenlänge k_n. Die S_n können hier für jedes n aus neuen Summanden bestehen.

Wir wollen die Folge der Verteilungen $(P_n^{S_n})_n$ untersuchen. Im Fall des Bernoulli-Experiments besagt der Grenzwertsatz von de Moivre und Laplace:

$$\lim_{n\to\infty} P\Big(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} < t\Big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-x^2/2)\, dx =: \Phi(t)$$

für $t \in \mathbb{R}$, d.h. die Verteilungsfunktionen $F_n(\cdot) := P\Big(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} < \cdot\Big)$ konvergieren punktweise gegen $\Phi(\cdot)$.

Zunächst klären wir den Zusammenhang zwischen Konvergenz in Verteilung bzw. schwacher Konvergenz und Konvergenz der zugehörigen Verteilungsfunktionen:

12.1 Satz *Es seien P und $(P_n)_n$ W-Maße auf $(\mathbb{R}, \mathcal{B})$ und F, $(F_n)_n$ die zugehörigen Verteilungsfunktionen, d.h. $F(x) := P((-\infty, x))$ und $F_n(x) := P_n((-\infty, x))$, $x \in \mathbb{R}$. Weiter sei W eine Teilmenge von $C_b(\mathbb{R})$ mit der folgenden Eigenschaft:*

Für alle $x < y$ existiert ein $f \in W$ mit $0 \leq f \leq 1$, $f(z) = 1$ für alle $z \leq x$, $f(z) = 0$ für alle $z \geq y$. Dann sind die folgenden Aussagen äquivalent:

(a)
$$\lim_{n\to\infty} \int f\, dP_n = \int f\, dP \quad \textit{für alle} \quad f \in C_b(\mathbb{R}).$$

(schwache Konvergenz)

(b)
$$\lim_{n\to\infty} \int f\, dP_n = \int f\, dP \quad \textit{für alle} \quad f \in W.$$

(c)
$$\lim_{n\to\infty} F_n(x) = F(x) \text{ \textit{für alle} } x, \text{ \textit{an denen} } F \text{ \textit{stetig ist.}}$$

12.2 Bemerkungen

(a) Der Satz von de Moivre und Laplace ist also eine Aussage über schwache Konvergenz bzw. Konvergenz in Verteilung, denn Φ ist überall stetig.

(b) Die Menge W nennt man *Konvergenz-determinierend*. Wir hatten in Satz 9.16 bereits gesehen, dass die Menge der gleichmäßig stetigen und beschränkten Funktionen eine Konvergenz-determinierende Menge ist.

(c) Gibt es die Funktionenmenge W? Es seien

$$W_0 := \{f \in C_b(\mathbb{R}) : f^{(k)} \in C_b(\mathbb{R}) \text{ für alle } k \in \mathbb{N}\}$$

und

$$f_0(t) := \begin{cases} 1, & \text{falls } t \leq 0, \\ 0, & \text{falls } t \geq 1, \\ \frac{\int_t^1 \exp\left(-\frac{1}{s(1-s)}\right) ds}{\int_0^1 \exp\left(-\frac{1}{s(1-s)}\right) ds}, & \text{falls } 0 < t < 1. \end{cases}$$

Dann ist f_0 wohldefiniert mit $0 \leq f_0 \leq 1$ und die Ableitungen $f_0^{(k)}$ existieren und sind in $C_b(\mathbb{R})$ für alle $k \in \mathbb{N}$. Also ist $f_0 \in W_0$. Für $x < y$ setzen wir nun $f(z) := f_0(\frac{z-x}{y-x})$. Dann hat f die gewünschten Eigenschaften.

Beweis von Satz 12.1: (b) $\Rightarrow$ (c): Zu $x \in \mathbb{R}$ und $\varepsilon > 0$ sei $f \in W$ so, dass

$$1_{(-\infty,x)} \leq f \leq 1_{(-\infty,x+\varepsilon)}.$$

Dann ist

$$\begin{aligned} \limsup_{n\to\infty} F_n(x) &= \limsup_{n\to\infty} P_n((-\infty,x)) \leq \limsup_{n\to\infty} \int f\, dP_n = \int f\, dP \\ &\leq P((-\infty, x+\varepsilon)) = F(x+\varepsilon) \end{aligned}$$

für alle $\varepsilon > 0$. Ganz analog folgt $\liminf_{n\to\infty} F_n(x) \geq F(x-\varepsilon)$. Daraus folgt $\lim_{n\to\infty} F_n(x) = F(x)$ für alle Stetigkeitsstellen von F.

(c) $\Rightarrow$ (a): (heißt in der Literatur der Satz von Helly und Bray) Sei D die Menge der Stetigkeitsstellen von F. Da F monoton wächst, ist $\mathbb{R} \setminus D$ abzählbar, D ist also dicht in $\mathbb{R}$. Sei $f \in C_b(\mathbb{R})$ und $\varrho := \sup_{x\in\mathbb{R}} |f(x)|$. Es existieren $\mu, \nu \in \mathbb{R}$ mit $F(\mu) < \varepsilon$ und $1 - F(\nu) < \varepsilon$, und da D dicht ist, existieren $\mu, \nu \in D$. Nun ist $f|_{[\mu,\nu]}$ gleichmäßig stetig, also existieren

$$\mu = \lambda_0 < \lambda_1 < \cdots < \lambda_{k-1} < \lambda_k = \nu \quad \text{mit } \lambda_0, \ldots, \lambda_k \in D\ , k \in \mathbb{N}_0\ ,$$

so dass $|f(x) - f(\lambda_{i-1})| < \varepsilon$ für $\lambda_{i-1} \le x \le \lambda_i$. Setze nun

$$g := \sum_{i=1}^{k} f(\lambda_{i-1}) 1_{[\lambda_{i-1}, \lambda_i)}.$$

Dann ist

$$\int g \, dP_n = \sum_{i=1}^{k} f(\lambda_{i-1}) (F_n(\lambda_i) - F_n(\lambda_{i-1})) \ .$$

Nach Voraussetzung gilt dann $\int g \, dP_n \to \int g \, dP$.

Sei nun $M := [\mu, \nu)$, $L = (-\infty, \mu)$ und $R = [\nu, \infty)$. Für $x \in L \cup R$ ist $|f(x) - g(x)| = |f(x)| \le \varrho$, für $x \in M$ ist $|f(x) - g(x)| < \varepsilon$, also

$$\begin{aligned} \Big| \int f \, dP - \int g \, dP \Big| &\le \int |f - g| \, dP \\ &\le \varrho P(L) + \varrho P(R) + \varepsilon P(M) \\ &= \varrho F(\mu) + \varrho (1 - F(\nu)) + \varepsilon P(M) \\ &\le (2\varrho + 1)\varepsilon \ . \end{aligned}$$

Analog folgt

$$\Big| \int f \, dP_n - \int g \, dP_n \Big| \le \varrho F_n(\mu) + \varrho(1 - F_n(\nu)) + \varepsilon \ .$$

Da μ, ν Stetigkeitstellen von F sind, folgt

$$\Big| \int f \, dP_n - \int g \, dP_n \Big| < (2\varrho + 1)\varepsilon$$

für n hinreichend groß und die Behauptung des Satzes folgt mit Hilfe der Dreiecks-Ungleichung. □

Wir wollen nun den folgenden Satz beweisen:

12.3 Satz (zentraler Grenzwertsatz) *Für jedes $n \in \mathbb{N}$ seien $X_{n1}, \dots, X_{nk_n}$ unabhängige Zufallsgrößen auf $(\Omega_n, \mathcal{A}_n, P_n)$ mit $\mathbb{E}(X_{nj}) = 0$ für alle $n \in \mathbb{N}$, $j = 1, \dots, k_n$ und endlichen Varianzen σ_{nk}^2, $n \ge 1$, $k = 1, \dots, k_n$. Es sei $s_n^2 := \sum_{k=1}^{k_n} \sigma_{nk}^2 > 0$ und*

$$\frac{1}{s_n^2} \sum_{k=1}^{k_n} \int_{\{|X_{nk}| \ge \varepsilon s_n\}} X_{nk}^2 \, dP_n \to 0 \quad \text{für} \ \ n \to \infty \ \ \text{für alle } \varepsilon > 0 \ , \tag{12.1}$$

dann gilt:

$$\lim_{n \to \infty} P_n \Big(\frac{S_n}{s_n} < y \Big) = \Phi(y)$$

für alle $y \in \mathbb{R}$.

Zunächst diskutieren wir die Bedingung (12.1), die einen eigenen Namen bekommt:

12.4 Definition Man sagt, dass die Folge von Zufallsgrößen $X_{n1}, \ldots, X_{nk_n}$, $n \in \mathbb{N}$, der LINDEBERG-*Bedingung* genügt, falls (12.1) gilt.

Die Bedingung (12.1) besagt, dass jeder Summand $\frac{X_{nk}}{s_n}$ von $S_n^* := \frac{S_n}{s_n}$ für große n nur einen kleinen Beitrag zur Gesamtsumme S_n^* liefert. Dies wird in dem folgenden Lemma präzisiert.

12.5 Lemma *Genügt* $X_{n1}, \ldots, X_{nk_n}$, $n \in \mathbb{N}$, *der* LINDEBERG-*Bedingung, so folgt:*

(a) $\max_{1 \le k \le k_n} \frac{|X_{nk}|}{s_n} \to 0$ stochastisch.

(b) $\lim_{n\to\infty} \frac{1}{s_n} \max_{1 \le k \le k_n} \sigma_{nk} = 0$ (FELLER-*Bedingung).*

Beweis: Zu (a): Es gilt mittels der MARKOV-Ungleichung

$$\begin{aligned} P_n\Big(\max_{1\le k\le k_n} \frac{|X_{nk}|}{s_n} \ge \varepsilon\Big) &= P_n\Big(\bigcup_{k=1}^{k_n}\Big\{\frac{|X_{nk}|}{s_n} \ge \varepsilon\Big\}\Big) \le \sum_{k=1}^{k_n} P_n(|X_{nk}| \ge \varepsilon s_n) \\ &\le \frac{1}{s_n^2\varepsilon^2}\sum_{k=1}^{k_n}\int_{\{|X_{nk}|\ge\varepsilon s_n\}} X_{nk}^2\, dP_n \ . \end{aligned}$$

Damit folgt mit Hilfe der LINDEBERG-Bedingung die Behauptung.

Zu (b): Wir bezeichnen zur Vereinfachung der Notation zu festem n mit $\mathbb{E}$ den Erwartungswert bezüglich P_n. Es gilt

$$\begin{aligned} \sigma_{nk}^2 &= \mathbb{E}(X_{nk}^2) = \int_{\{|X_{nk}|\ge\varepsilon s_n\}} X_{nk}^2\, dP_n + \int_{\{|X_{nk}|<\varepsilon s_n\}} X_{nk}^2\, dP_n \\ &\le \int_{\{|X_{nk}|\ge\varepsilon s_n\}} X_{nk}^2\, dP_n + \varepsilon^2 s_n^2, \end{aligned}$$

und somit

$$\limsup_{n\to\infty} \max_{1\le k\le k_n} \frac{\sigma_{nk}^2}{s_n^2} \le \varepsilon^2 \text{ für alle } \varepsilon > 0 \ .$$

□

12.6 Lemma *Sind die Zufallsgrößen* $\big(X_{nk}/\sigma_{nk}\big)_{n\ge1}^{k=1,\ldots,k_n}$ *zusätzlich identisch verteilt, so folgt aus der* FELLER*schen Bedingung die* LINDEBERG-*Bedingung.*

Beweis: Setze $\varrho_n := \max_{1\le k\le k_n} \frac{\sigma_{nk}}{s_n}$, dann gilt

$$\begin{aligned}\int_{\{|X_{nk}|\ge\varepsilon s_n\}} X_{nk}^2\,dP_n &= \sigma_{nk}^2 \int_{\left\{\frac{|X_{nk}|}{\sigma_{nk}}\ge\frac{\varepsilon}{\sigma_{nk}/s_n}\right\}} \left(\frac{X_{nk}}{\sigma_{nk}}\right)^2 dP_n \\ &\le \sigma_{nk}^2 \int_{\left\{\frac{|X_{nk}|}{\sigma_{nk}}\ge\frac{\varepsilon}{\varrho_n}\right\}} \left(\frac{X_{nk}}{\sigma_{nk}}\right)^2 dP_n \\ &= \sigma_{nk}^2 \int 1_{[\varepsilon/\varrho_n,\infty)}(|x|)\; x^2\, P_n^{\frac{X_{nk}}{\sigma_{nk}}}(dx).\end{aligned}$$

Die rechte Seite konvergiert für $n\to\infty$ gegen Null, denn die Verteilung $P_n^{\frac{X_{nk}}{\sigma_{nk}}}$ ist unabhängig von n. □

12.7 Korollar *Für jedes $n \ge 1$ seien $X_{n1},\dots,X_{nk_n}$ unabhängig und die $(X_{nk})_{n\ge1}^{k=1,\dots,k_n}$ seien identisch verteilt mit $\mathbb{E}(X_{11}) = 0$, $\sigma_{11}^2 = \mathbb{E}(X_{11}^2) < \infty$ und $s_n^2 = \sum_{k=1}^{k_n}\sigma_{nk}^2 > 0$. Dann gilt*

$$\lim_{n\to\infty} P_n\Big(\frac{\sum_{k=1}^{k_n} X_{nk}}{s_n} \le y\Big) = \Phi(y) \quad \textit{für alle} \quad y\in\mathbb{R}\ .$$

Erneut bezeichnet hier $\mathbb{E}$ den Erwartungswert bezüglich P_n zu festem n.

Beweis: Nach Lemma 12.6 genügt es, $\varrho_n = \frac{1}{s_n}\max_{1\le k\le k_n}\sigma_{nk} \to 0$ für $n\to\infty$ zu zeigen. Hier ist nun $\varrho_n = \frac{\sigma_{nk}}{\sqrt{k_n\sigma_{nk}^2}} \to 0$ für $n\to\infty$. □

12.8 Korollar *Sei in der Situation von Korollar 12.7 $X_{nk} = X_k$, $1\le k\le k_n = n$, $n\ge 1$, also eine Folge von Zufallsgrößen, definiert auf einem W-Raum $(\Omega,\mathcal{A},P)$. Dann gilt*

$$\liminf_{n\to\infty}\frac{S_n}{\sqrt{n}} = -\infty \ \textit{und}\ \limsup_{n\to\infty}\frac{S_n}{\sqrt{n}} = +\infty$$

P-fast sicher.

(Nachtrag zur Diskussion vor Satz 11.11)

Beweis: Ohne Einschränkung sei $\mathrm{Var}(X_{n1}) = 1$. Wir wissen, dass $\liminf_{n\to\infty} S_n/\sqrt{n}$ und $\limsup_{n\to\infty} S_n/\sqrt{n}$ terminale Funktionen sind, also konstant α (bzw. $-\alpha$) mit $\alpha\in\bar{\mathbb{R}}$, siehe Korollar 10.20. Angenommen $-\alpha > -\infty$. Dann gilt für $t\in(-\infty,-\alpha)$

$$\begin{aligned} 0 &= P\Big(\liminf_{n\to\infty} S_n/\sqrt{n} < -\alpha\Big) = P\Big(\bigcap_{n\ge1}\Big\{\inf_{m\ge n} S_m/\sqrt{m} < -\alpha\Big\}\Big) \\ &\ge \lim_{n\to\infty} P\Big(\inf_{m\ge n} S_m/\sqrt{m} < t\Big) \ge \lim_{n\to\infty} P\Big(S_n/\sqrt{n} < t\Big) = \Phi(t) > 0\ ,\end{aligned}$$

also ein Widerspruch. □

Wir kommen nun zum Beweis von Satz 12.3. Nach Satz 12.1 wollen wir zeigen:

$$\lim_{n\to\infty} \int f\left(\frac{S_n}{s_n}\right) dP_n = \mathbb{E}\big(f(N)\big) \text{ für alle } f \in W \ ,$$

wobei N eine standardnormalverteilte Zufallsgröße bezeichne.

Die Idee, die auf J.W. Lindeberg zurückgeht, wird sein, auf der linken Seite die X_{nk} durch $N(0,\sigma_{nk}^2)$-verteilte Zufallsgrößen zu ersetzen, und zwar sukzessive. Man erhält dann eine Folge der Form

$$\begin{aligned}
&\int f\Big(\frac{X_{n1}+\cdots+X_{nk_n}}{s_n}\Big)\, dP_n \ ,\\
&\int f\Big(\frac{X_{n1}+\cdots+X_{nk_n-1}+Y_{nk_n}}{s_n}\Big)\, dP_n \ ,\ldots,\\
&\int f\Big(\frac{X_{n1}+Y_{n2}+\cdots+Y_{nk_n}}{s_n}\Big)\, dP_n \ ,\\
&\int f\Big(\frac{Y_{n1}+\cdots+Y_{nk_n}}{s_n}\Big)\, dP_n \ .
\end{aligned}$$

Wir zeigen dann, dass jedes Glied dieser Folge für große n so nahe beim Nächsten liegt, dass sogar das erste und letzte Glied nahe zusammen liegen! Wenn die Y_{nk}, $k=1,\ldots,k_n$, unabhängig gewählt werden, ist das letzte Glied gleich $\mathbb{E}\big(f(N)\big)$. Diesen Sachverhalt kennt man aus der Vorlesung des dritten Semesters:

12.9 Satz *Seien X_1 und X_2 unabhängige Zufallsgrößen und $P^{X_1} = N(\mu,\sigma^2)$, $P^{X_2} = N(\nu,\tau^2)$, dann ist $P^{X_1} * P^{X_2} = N(\mu+\nu, \sigma^2+\tau^2)$.*

Beweis: Übung, indem man die Faltungsdichte in Satz 10.24 ausrechnet. Wir werden später noch einen elementaren Beweis kennenlernen.

Die Aussage und die Voraussetzung des Satzes 12.3 sind nur abhängig von $P_n^{(X_{n1},\ldots,X_{nk_n})}$. Wir werden nun die gegebene Folge $X_{n1},\ldots,X_{nk_n}$ von Zufallsgrößen auf $(\Omega_n,\mathcal{A}_n,P_n)$ geeignet durch eine unabhängige Folge $\tilde{X}_{n1},\ldots,\tilde{X}_{nk_n},Y_{n1},\ldots,Y_{nk_n}$ von Zufallsgrößen auf einem W-Raum $(\tilde{\Omega}_n,\tilde{\mathcal{A}}_n,\tilde{P}_n)$ ersetzen: Dazu sei für $n\geq 1$ $(\tilde{\Omega}_n,\tilde{\mathcal{A}}_n,\tilde{P}_n)$ ein W-Raum und

$$\begin{aligned}
\mathcal{N} := \Big\{ \tilde{X} = \big((\tilde{X}_{n1},\ldots,\tilde{X}_{nk_n})\big)_{n\geq 1} \Big| &\forall n \geq 1 : \tilde{X}_{n1},\ldots,\tilde{X}_{nk_n} \in \mathcal{L}^2(\tilde{P}_n),\\
&\text{unabhängig, zentriert und } s_n^2 > 0,\ \tilde{X} \text{ genüge der Lindeberg-}\\
&\text{Bedingung} \Big\} \ ,
\end{aligned}$$

sowie

$$\mathcal{N}_0 := \Big\{ Y \in \mathcal{N} : \lim_{n\to\infty} \tilde{P}_n\Big(\frac{\sum_{k=1}^{k_n} Y_{nk}}{s_n} \leq y\Big) = \Phi(y) \text{ für alle } y \in \mathbb{R} \Big\} .$$

Wir zeigen dass W-Räume $(\tilde{\Omega}_n, \tilde{\mathcal{A}}_n, \tilde{P}_n)$, $n \geq 1$, und Zufallsgrößen $\tilde{X}_{n1}, \ldots, \tilde{X}_{nk_n}, Y_{n1}, \ldots, Y_{nk_n}$ darauf existieren mit:

(a) $\tilde{P}_n^{\tilde{X}_{nk}} = P_n^{X_{nk}}$, $n \geq 1, k = 1, \ldots, k_n$

(b) $\mathrm{Var}(\tilde{X}_{nk}) = \mathrm{Var}(Y_{nk})$, $n \geq 1, k = 1, \ldots, k_n$

(c) $\tilde{X}_{n1}, \ldots, \tilde{X}_{nk_n}, Y_{n1}, \ldots, Y_{nk_n}$ sind unabhängig

(d) $Y = \big((Y_{n1}, \ldots, Y_{nk_n})\big)_{n\geq 1} \in \mathcal{N}_0$.

In (b) sind die Varianz bezüglich $\tilde{P}_n$ gemeint. Präziser müssen wir $\mathrm{Var}_{\tilde{P}_n}(\tilde{X}_{nk})$ schreiben. Im Fall der obigen Konstruktion gilt $\tilde{X} = \big((\tilde{X}_{n1}, \ldots, \tilde{X}_{nk_n})\big)_{n\geq 1} \in \mathcal{N}$, denn

$$\mathrm{Var}(\tilde{X}_{nk}) = \mathrm{Var}(X_{nk}) = \sigma_{nk}^2 \text{ , also } s_n^2 > 0,$$

und

$$\begin{aligned}
\int_{\{|X_{nk}|\geq \varepsilon s_n\}} X_{nk}^2 \, dP_n &= \int 1_{[\varepsilon s_n, +\infty)}(|X_{nk}|)|X_{nk}|^2 \, dP_n \\
&= \int 1_{[\varepsilon s_n, +\infty)}(|x|)|x|^2 \, dP_n^{X_{nk}}(x) \\
&= \int 1_{[\varepsilon s_n, +\infty)}(|x|)|x|^2 \, d\tilde{P}_n^{\tilde{X}_{nk}}(x) \ ,
\end{aligned}$$

also erfüllt $\tilde{X}$ die LINDEBERG-Bedingung.

Das sogenannte *Invarianzprinzip* führt uns sogar zu $\tilde{X} \in \mathcal{N}_0$, d.h. $\lim_{n\to\infty} \tilde{P}_n\Big(\sum_{k=1}^{k_n} \tilde{X}_{nk}/s_n \leq y\Big) = \Phi(y)$, aber $\tilde{P}_n\Big(\sum_{k=1}^{k_n} \tilde{X}_{nk}/s_n \leq y\Big) = P_n\Big(\sum_{k=1}^{k_n} X_{nk}/s_n \leq y\Big)$, denn

$$\begin{aligned}
P_n^{(X_{n1}, \ldots, X_{nk_n})} &= P_n^{X_{n1}} \otimes \cdots \otimes P_n^{X_{nk_n}} \\
&= \tilde{P}_n^{\tilde{X}_{n1}} \otimes \cdots \otimes \tilde{P}_n^{\tilde{X}_{nk_n}} = \tilde{P}_n^{(\tilde{X}_{n1}, \ldots, \tilde{X}_{nk_n})} \ ,
\end{aligned}$$

womit alles gezeigt ist.

12.10 Satz (Invarianzprinzip) *Es seien $\tilde{X} \in \mathcal{N}$ und $Y \in \mathcal{N}_0$ mit*

(a) $\mathrm{Var}(\tilde{X}_{nk}) = \mathrm{Var}(Y_{nk})$, $n \geq 1$, $k = 1, \ldots, k_n$

(b) $\tilde{X}_{n1}, \dots, \tilde{X}_{nk_n}, Y_{n1}, \dots, Y_{nk_n}$ *sind unabhängig für jedes* $n \geq 1$.
Dann gilt $\tilde{X} \in \mathcal{N}_0$.

Wir müssen dieses Prinzip beweisen, zeigen aber zuvor, dass obige Konstruktion mit (a), …, (d) möglich ist. Dazu seien

$$\tilde{\Omega}_n := \mathbb{R}^{2k_n} \ , \ \tilde{\mathcal{A}}_n := \mathcal{B}^{2k_n} \text{ und } \tilde{P}_n := P_n^{X_{n1}} \otimes \cdots \otimes P_n^{X_{nk_n}} \otimes Q^{n1} \otimes \cdots \otimes Q^{nk_n}$$

mit $Q^{nk} = N(0, \sigma_{nk}^2)$, $k = 1, \dots, k_n$. Weiter seien

$$\tilde{X}_{nk} : \tilde{\Omega}_n \to \mathbb{R} \text{ mit } \tilde{X}_{nk}(x_{n1}, \dots, x_{nk_n}, y_{n1}, \dots, y_{nk_n}) = x_{nk}$$

und

$$\tilde{Y}_{nk} : \tilde{\Omega}_n \to \mathbb{R} \text{ mit } \tilde{Y}_{nk}(x_{n1}, \dots, x_{nk_n}, y_{n1}, \dots, y_{nk_n}) = y_{nk}$$

für $k = 1, \dots, k_n$. Die Abbildungen sind Projektionen. Dann gilt (a)-(c) nach Konstruktion. Für (d) müssen wir $Y \in \mathcal{N}$ zeigen, was nach Definition von $\mathcal{N}$ sich darauf beschränkt, die Lindeberg-Bedingung nachzurechnen. Nun ist $\frac{Y_{nk}}{\sigma_{nk}}$ $N(0,1)$-verteilt für alle $k = 1, \dots, k_n$, also zeigen wir gemäß Lemma 12.6

$$\lim_{n\to\infty} \frac{1}{s_n} \max_{1\leq k\leq k_n} \sigma_{nk} = 0.$$

Da $\sigma_{nk}^2 = \operatorname{Var}(X_{nk})$ und X nach Voraussetzung die Lindeberg-Bedingung erfüllt, folgt dies aber unmittelbar aus Lemma 12.5(b). Also ist $Y \in \mathcal{N}$. Mit Satz 12.9 ist weiter

$$\tilde{P}_n\Big(\frac{\sum_{k=1}^{k_n} Y_{nk}}{s_n}\Big) = \Phi(y) \qquad \text{für alle } n \in \mathbb{N}$$

und für alle $y \in \mathbb{R}$, also ist $Y \in \mathcal{N}_0$.

Es bleibt also, das Invarianzprinzip zu beweisen:

Beweis von Satz 12.10: 1. Schritt: Es sei $f \in W_0$ und für $x, h \in \mathbb{R}$

$$g(x,h) := \begin{cases} \frac{2}{h^2}\big(f(x+h) - f(x) - h\,f'(x)\big) - f''(x), & \text{falls } h \neq 0, \\ 0, & \text{falls } h = 0, \end{cases}$$

also

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}(f''(x) + g(x,h)) \ , x, h \in \mathbb{R} \ . \tag{12.2}$$

Nach Taylor ist

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x + \theta h)$$

für ein $0 < \theta < 1$, welches von x und h abhängt. Damit folgt $|g(x,h)| \leq 2\sup_{x\in\mathbb{R}}|f''(x)| =: A$ für alle $x, h \in \mathbb{R}$ (A existiert, da $f \in W_0$). Nach TAYLOR gilt auch

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x+\theta' h)$$

mit $0 < \theta' < 1$, also $|g(x,h)| \leq |h|\frac{1}{3}\sup_{x\in\mathbb{R}}|f'''(x)| =: |h|B$ (B existiert, da $f \in W_0$). Insgesamt haben wir (12.2) mit

$$|g(x,h)| \leq \min\{A, |h|B\} =: d(h),\ x,h \in \mathbb{R}\ .$$

2. Schritt: Wir setzen nun $\tilde{T}_n := \frac{1}{s_n}\sum_{k=1}^{k_n}\tilde{X}_{nk}$, $\hat{T}_n := \frac{1}{s_n}\sum_{k=1}^{k_n}Y_{nk}$, sowie $\tilde{P}_n^{\tilde{T}_n} =: \tilde{Q}_n$ und $\tilde{P}_n^{\hat{T}_n} =: \hat{Q}_n$. Zu zeigen:

$$\lim_{n\to\infty}\int f\,d\tilde{Q}_n = \mathbb{E}\big(f(N)\big) \quad \text{für alle } f \in W_0\ .$$

Also ist zu zeigen:

$$\lim_{n\to\infty}\Big|\int f\,d\tilde{Q}_n - \int f\,d\hat{Q}_n\Big| = 0 \ \ \text{für alle } f \in W_0\ .$$

Es gilt $\int f\,d\tilde{Q}_n - \int f\,d\hat{Q}_n = \int f\circ\tilde{T}_n\,d\tilde{P}_n - \int f\circ\hat{T}_n\,d\tilde{P}_n$. Sei nun

$$U_{nk} := \frac{1}{s_n}(\tilde{X}_{n1} + \cdots + \tilde{X}_{n,k-1} + Y_{n,k+1} + \cdots + Y_{nk_n})$$

für jedes $k = 1, \ldots, k_n$, also

$$f\circ\tilde{T}_n - f\circ\hat{T}_n = \sum_{k=1}^{k_n}\Big(f\big(U_{nk} + \frac{\tilde{X}_{nk}}{s_n}\big) - f\big(U_{nk} + \frac{Y_{nk}}{s_n}\big)\Big)\ .$$

Es gilt

$$\begin{aligned}
&\mathbb{E}_{\tilde{P}_n}\Big(f\big(U_{nk} + \frac{\tilde{X}_{nk}}{s_n}\big)\Big)\\
&= \mathbb{E}_{\tilde{P}_n}\Big(f\big(U_{nk}\big)\Big) + \mathbb{E}_{\tilde{P}_n}\Big(\frac{\tilde{X}_{nk}}{s_n}\,f'(U_{nk})\Big) + \mathbb{E}_{\tilde{P}_n}\Big(\frac{\tilde{X}_{nk}^2}{2s_n^2}f''\big(U_{nk}\big)\Big)\\
&+ \mathbb{E}_{\tilde{P}_n}\Big(\frac{\tilde{X}_{nk}^2}{2s_n^2}g\big(U_{nk}, \frac{\tilde{X}_{nk}}{s_n}\big)\Big)\\
&= \mathbb{E}_{\tilde{P}_n}\Big(f\big(U_{nk}\big)\Big) + \frac{1}{2s_n^2}\operatorname{Var}_{\tilde{P}_n}(\tilde{X}_{nk})\mathbb{E}_{\tilde{P}_n}\big(f''(U_{nk})\big) + \mathbb{E}_{\tilde{P}_n}\Big(\frac{\tilde{X}_{nk}^2}{2s_n^2}g\big(U_{nk}, \frac{\tilde{X}_{nk}}{s_n}\big)\Big)\ ,
\end{aligned}$$

denn die $\tilde{X}_{nk}$'s sind zentriert und wir haben Satz 10.21 verwendet. Für $\mathbb{E}_{\tilde{P}_n}\Big(f\big(U_{nk}+\frac{Y_{nk}}{s_n}\big)\Big)$ folgt die analoge Identität, wenn wir $\tilde{X}_{nk}$ durch Y_{nk} ersetzen. Mit $\mathrm{Var}_{\tilde{P}_n}(\tilde{X}_{nk}) = \mathrm{Var}_{\tilde{P}_n}(Y_{nk})$ folgt

$$\int f\,d\tilde{Q}_n - \int f\,d\hat{Q}_n = \frac{1}{2s_n^2}\sum_{k=1}^{k_n}\Big(\mathbb{E}_{\tilde{P}_n}\Big(\tilde{X}_{nk}^2 g\big(U_{nk}, \frac{\tilde{X}_{nk}}{s_n}\big)\Big) - \mathbb{E}_{\tilde{P}_n}\Big(Y_{nk}^2 g\big(U_{nk}, \frac{\tilde{X}_{nk}}{s_n}\big)\Big)\Big),$$

also

$$\Big|\int f\,d\tilde{Q}_n - \int f\,d\hat{Q}_n\Big| \le \frac{1}{2s_n^2}\sum_{k=1}^{k_n}\Big(\mathbb{E}_{\tilde{P}_n}\Big(\tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\Big) + \mathbb{E}_{\tilde{P}_n}\Big(Y_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\Big)\Big) .$$

Wenn wir nun zeigen, dass die rechte Seite gegen Null konvergiert für $n \to \infty$, sind wir fertig. Es gilt

$$\begin{aligned}
\mathbb{E}_{\tilde{P}_n}\Big(\tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\Big) &= \int_{\{|\tilde{X}_{nk}|\ge\delta s_n\}} \tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\,d\tilde{P}_n + \int_{\{|\tilde{X}_{nk}|<\delta s_n\}} \tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\,d\tilde{P}_n \\
&\le A\int_{\{|\tilde{X}_{nk}|\ge\delta s_n\}} \tilde{X}_{nk}^2\,d\tilde{P}_n + B\delta\,\mathrm{Var}_{\tilde{P}_n}(\tilde{X}_{nk}) ,
\end{aligned}$$

also für alle $n \in \mathbb{N}$ und $\delta > 0$.

$$\frac{1}{2s_n^2}\sum_{k=1}^{k_n}\mathbb{E}_{\tilde{P}_n}\Big(\tilde{X}_{nk}^2 d\big(\frac{\tilde{X}_{nk}}{s_n}\big)\Big) \le \frac{A}{2s_n^2}\sum_{k=1}^{k_n}\int_{\{|\tilde{X}_{nk}|\ge\delta s_n\}} \tilde{X}_{nk}^2\,d\tilde{P}_n + \frac{B\delta}{2s_n^2}s_n^2 .$$

Die rechte Seite wird gemäß der Lindeberg-Bedingung klein in $n \in \mathbb{N}$. Analog folgt

$$\lim_{n\to\infty}\frac{1}{2s_n^2}\sum_{k=1}^{k_n}\mathbb{E}_{\tilde{P}_n}\Big(Y_{nk}^2 d\big(\frac{Y_{nk}}{s_n}\big)\Big) = 0 .$$

□

KAPITEL 13

Bedingte Erwartungswerte

Bisher haben wir ausführlich Eigenschaften unabhängiger Zufallsgrößen untersucht. Wir wollen nun beginnen, stochastische Abhängigkeiten in einem stochastischen Modell mathematisch präzise zu beschreiben. Der Begriff des bedingten Erwartungswerts ist zentral hierzu.

Gegeben sei fortan ein W-Raum $(\Omega, \mathcal{A}, P)$. Zu $A, B \in \mathcal{A}$ mit $P(B) > 0$ heißt

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

(elementare) *bedingte Wahrscheinlichkeit* von A unter (der Hypothese) B. Die Mengenfunktion $P(\cdot|B) : \mathcal{A} \to [0,1]$ bildet ein W-Maß auf $(\Omega, \mathcal{A})$. Seien X, Y zwei Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$, Y sei diskret. Dann können wir

$$P(X \in B|Y = y) = P(\{X \in B\}|\{Y = y\}) = \frac{P(X \in B, Y = y)}{P(Y = y)}$$

für alle $B \in \mathcal{B}$ und y mit $P(Y = y) > 0$ definieren. Hierbei bezeichnet $\mathcal{B}$ die σ-Algebra der Borelschen Mengen. Für jedes y ist $P(X \in \cdot|Y = y)$ ein W-Maß auf $(\mathbb{R}, \mathcal{B})$, welches häufig *bedingte Verteilung* von X, gegeben $Y = y$, genannt wird, in Zeichen $P^{X|Y=y}$. Der zugehörige Erwartungswert

$$\mathbb{E}(X|Y = y) := \int x \, P^{X|Y=y}(dx) = \frac{1}{P(Y = y)} \int_{\{Y=y\}} X \, dP \,, \tag{13.1}$$

sofern dieser existiert, heißt *bedingter Erwartungswert* von X gegeben $Y = y$. Die zweite Gleichheit folgt mittels eines Funktions-Erweiterungsarguments (man zeige die Gleichheit zunächst für Indikatorabbildungen, dann für nichtnegative messbare numerische Abbildungen und schliesslich mittels Zerlegung in Positiv- und Negativteil für Zufallsgrößen X). Bei diskretem Y sind bedingte Wahrscheinlichkeit und bedingter Erwartungswert, gegeben $Y = y$, für P^Y-fast alle y definiert. Was ist aber zu tun, wenn Y keine diskrete Verteilung besitzt? Nun kann sogar $P(Y = y) = 0$ für alle $y \in \mathbb{R}$ gelten. Wir geben dazu, also auf der Suche nach einer allgemeinen Definition, eine *andere* Interpretation bedingter Wahrscheinlichkeit bzw. bedingter Erwartungswerte.

Sei zunächst wieder Y diskret und $|X| \le c$. Dann folgt $|\mathbb{E}(X|Y = y)| \le c$. Also bildet

$$f(y) := \begin{cases} \mathbb{E}(X|Y = y), & \text{falls } P(Y = y) > 0, \\ 0, & \text{sonst} \end{cases}$$

eine durch c beschränkte, messbare Funktion. Folglich ist $f(Y) =: \mathbb{E}(X|Y)$ eine beschränkte $\sigma(Y)$-messbare Zufallsgröße. Sei X eine Variable, deren Wert wir nicht beobachten können. Stattdessen erhalten wir den Wert von Y und suchen nach einer Approximation $g(Y)$, eine $\sigma(Y)$-messbare Variable, die möglichst gut X approximiert (wobei wir mittels einer Metrik klären müssen, was hier möglichst gut bedeuten soll). Bei diskretem Y soll das Ergebnis $f(Y) = \mathbb{E}(X|Y)$ sein. Gibt es eine Metrik, bezüglich der $f(Y)$ die beste Approximation liefert?

Wir beobachten zunächst: Für jedes $A = \{Y \in B\}$ gilt

$$\begin{aligned}\int_A \mathbb{E}(X|Y)\,dP &= \int_B \mathbb{E}(X|Y=y)\,P^Y(dy) = \sum_{y\in B} \mathbb{E}(X|Y=y)P(Y=y) \\ &= \sum_{y\in B} \Big(\frac{1}{P(Y=y)} \int_{\{Y=y\}} X\,dP)\Big)P(Y=y) = \int_{\{Y\in B\}} X\,dP.\end{aligned}$$

Hierbei haben wir die Transformationsformel sowie (13.1) verwendet. Es gilt also:

(a) $\mathbb{E}(X|Y)$ ist $\sigma(Y)$-messbar.

(b) $\int_A \mathbb{E}(X|Y)\,dP = \int_A X\,dP$ für alle $A \in \sigma(Y)$.

Durch (a) und (b) ist $\mathbb{E}(X|Y)$ fast sicher eindeutig bestimmt, denn sei Z eine $\sigma(Y)$-messbare Zufallsgröße mit $\int_A Z\,dP = \int_A X\,dP$ für alle $A \in \sigma(Y)$, so gilt

$$\int_A (\mathbb{E}(X|Y) - Z)\,dP = 0 \ \text{ für alle } A \in \sigma(Y)\ .$$

Da $\mathbb{E}(X|Y) - Z$ $\sigma(Y)$-messbar, folgt $\mathbb{E}(X|Y) - Z = 0$.

(b) können wir auch so schreiben:

$$\text{(b')}\ \mathbb{E}(1_A \mathbb{E}(X|Y)) = \mathbb{E}(1_A X) \text{ für alle } A \in \sigma(Y)\ .$$

Nun sind -im Moment nach Voraussetzung- $\mathbb{E}(X|Y)$ und X beschränkt, also existiert $\mathbb{E}\big(Z\,\mathbb{E}(X|Y)\big)$ und $\mathbb{E}(ZX)$ für beliebige integrierbare Zufallsgrößen Z. Mit Hilfe des Funktions-Erweiterungsarguments (siehe oben) liefert (b') dann

$$\text{(c)}\ \mathbb{E}\big(Z\,\mathbb{E}(X|Y)\big) = \mathbb{E}(ZX), \text{ also } \mathbb{E}\Big(Z\big(X - \mathbb{E}(X|Y)\big)\Big) = 0$$

für alle $\sigma(Y)$-messbaren und integrierbaren Zufallsgrößen Z.

Ist Z quadrat-integrierbar und $\sigma(Y)$-messbar, folgt

$$\begin{aligned}\mathbb{E}(X-Z)^2 &= \mathbb{E}\big(X - \mathbb{E}(X|Y)\big)^2 + \mathbb{E}\big(\mathbb{E}(X|Y) - Z\big)^2 \\ &\quad + 2\mathbb{E}\Big(\big(\mathbb{E}(X|Y) - Z\big)\big(X - \mathbb{E}(X|Y)\big)\Big) \\ &= \mathbb{E}\big(X - \mathbb{E}(X|Y)\big)^2 + \mathbb{E}\big(\mathbb{E}(X|Y) - Z\big)^2 \geq \mathbb{E}\big(X - \mathbb{E}(X|Y)\big)^2\ ,\end{aligned}$$

denn $\mathbb{E}(X|Y) - Z$ ist $\sigma(Y)$-messbar. Also minimiert $\mathbb{E}(X|Y)$ den $\mathcal{L}^2$-Pseudo-Abstand zu X unter allen $\sigma(Y)$-messbaren Z!

Wegen der Vollständigkeit von $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ bildet dieser Raum mit $\langle X, Y\rangle := \mathbb{E}(XY)$ einen Hilbertraum (siehe Satz 7.12), womit orthogonale Projektionen existieren, die bei der Suche nach besten Approximationen bzgl. der Norm die natürlichen Kandidaten darstellen.

Es sei nun $\mathcal{F} \subset \mathcal{A}$ eine σ-Algebra und $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Obiges (a) und (b) übertragen wir nun zu

(I) Die Zufallsgröße $\mathbb{E}(X|\mathcal{F})$ ist $\mathcal{F}$-messbar.

(II) $\displaystyle\int_A \mathbb{E}(X|\mathcal{F})\, dP = \int_A X\, dP$ für alle $A \in \mathcal{F}$.

Es gilt $\mathcal{L}^2(\Omega, \mathcal{F}, P) \subset \mathcal{L}^2(\Omega, \mathcal{A}, P)$.

Die *orthogonale Projektion* $\mathbb{P}_{\mathcal{F}}X$ von $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ ist durch

$$||X - \mathbb{P}_{\mathcal{F}}X||_2 = \min_{Z\in\mathcal{L}^2(\Omega,\mathcal{F},P)} ||X - Z||_2$$

fast sicher eindeutig bestimmt und auch durch

$$\langle X - \mathbb{P}_{\mathcal{F}}X, Z\rangle = \mathbb{E}\big((X - \mathbb{P}_{\mathcal{F}}X)Z\big) = 0 \tag{13.2}$$

für alle $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ charakterisiert. Man nennt $\mathbb{P}_{\mathcal{F}}X$ *beste Approximation* von X in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$.

13.1 Satz und Definition Seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $\mathcal{F}$ eine Unter-σ-Algebra von $\mathcal{A}$ und $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Bezeichnet dann $\mathbb{E}(X|\mathcal{F})$ die P-f.s. eindeutige beste Approximation von X in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, so erfüllt diese die Bedingung (I) und (II). $\mathbb{E}(X|\mathcal{F})$ heißt *bedingter Erwartungswert* oder *bedingte Erwartung* von X gegeben $\mathcal{F}$.

Beweis: Mit (13.2) gilt für alle $\mathcal{A} \in \mathcal{F}$ (womit $1_A \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$)

$$\int_A \big(X - \mathbb{E}(X|\mathcal{F})\big)\, dP = \langle X - \mathbb{E}(X|\mathcal{F}), 1_A\rangle = 0. \qquad \Box$$

Der bedingte Erwartungswert ist außerhalb von P-Nullmengen $\in \mathcal{F}$ eindeutig bestimmt. Man bezeichnet jede zulässige Wahl auch als *Version* von $\mathbb{E}(X|\mathcal{F})$. Wir setzen $\mathbb{E}(X|Y) := \mathbb{E}\big(X|\sigma(Y)\big)$.

13.2 Lemma *Seien $(\Omega, \mathcal{A}, P)$ ein W-Raum und $\mathcal{F}$ eine Unter-σ-Algebra von $\mathcal{A}$ und $X, Y, X_1, X_2, \ldots \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Dann gilt*

(a) $\mathbb{E}(\alpha X + \beta Y|\mathcal{F}) = \alpha\mathbb{E}(X|\mathcal{F}) + \beta\mathbb{E}(Y|\mathcal{F})$ *P-f.s. für alle* $\alpha, \beta \in \mathbb{R}$.

(b) $X \le Y$ *P-f.s.* $\Rightarrow \mathbb{E}(X|\mathcal{F}) \le \mathbb{E}(Y|\mathcal{F})$ *P-f.s.*

(c) $X_n \nearrow X$ *P-f.s.* $\Rightarrow \mathbb{E}(X_n|\mathcal{F}) \nearrow \mathbb{E}(X|\mathcal{F})$ *P-f.s.*

Beweis:

(a) Es gilt

$$\begin{aligned} 0 &= \alpha \langle X - \mathbb{E}(X|\mathcal{F}), Z\rangle + \beta \langle Y - \mathbb{E}(Y|\mathcal{F}), Z\rangle \\ &= \Big\langle (\alpha X + \beta Y) - \big(\alpha\mathbb{E}(X|\mathcal{F}) + \beta\mathbb{E}(Y|\mathcal{F})\big), Z\Big\rangle \end{aligned}$$

für alle $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Die P-f.s. Eindeutigkeit des bedingten Erwartungswertes liefert die Behauptung.

(b) $A := \{\mathbb{E}(X|\mathcal{F}) > \mathbb{E}(Y|\mathcal{F})\} \in \mathcal{F}$. Zu zeigen: $P(A) = 0$. Mit (II) gilt nach Definition von A

$$0 \le \int_A \big(\mathbb{E}(X|\mathcal{F}) - \mathbb{E}(Y|\mathcal{F})\big)\, dP = \int_A (X - Y)\, dP \le 0 \ ,$$

da $X \le Y$ P-f.s. Damit folgt wegen der strikten Positivität von $\mathbb{E}(X|\mathcal{F}) - \mathbb{E}(Y|\mathcal{F})$ auf A, dass $P(A) = 0$.

(c) Nach (b) existiert ein $\mathcal{F}$-messbares W mit $\mathbb{E}(X_n|\mathcal{F}) \nearrow W$ P-f.s. Da $\mathbb{E}(X_1|\mathcal{F}) \le W \le \mathbb{E}(X|\mathcal{F})$ P-f.s., folgt $W \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Es gilt

$$\begin{aligned} \langle X - W, Z\rangle &= \langle X - X_n, Z\rangle + \langle X_n - \mathbb{E}(X_n|\mathcal{F}), Z\rangle + \langle \mathbb{E}(X_n|\mathcal{F}) - W, Z\rangle \\ &= \langle X - X_n, Z\rangle + \langle \mathbb{E}(X_n|\mathcal{F}) - W, Z\rangle \\ &\le \big(||X - X_n||_2 + ||W - \mathbb{E}(X_n|\mathcal{F})||_2\big)||Z||_2 \end{aligned}$$

nach Cauchy-Schwarz. Da $0 \le X - X_n \le X - X_1$ und $0 \le W - \mathbb{E}(X_n|\mathcal{F}) \le W - \mathbb{E}(X_1|\mathcal{F})$ P-f.s., liefert der Satz von der dominierten Konvergenz, dass der letzte Ausdruck für $n \to \infty$ gegen 0 strebt. Wegen der P-f.s. Eindeutigkeit des bedingten Erwartungswertes muss $W = \mathbb{E}(X|\mathcal{F})$ P-f.s. gelten. $\square$

Wir wollen den Begriff des bedingten Erwartungswertes auf quasi-integrierbare Zufallsvariablen ausdehnen. Sei zunächst $X \ge 0$ P-f.s. und $X_n := X1_{\{X \le n\}}$, $n \ge 1$. Dann ist $X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ und mit Lemma 13.2 (b) existiert zu jeder Unter-σ-Algebra $\mathcal{F} \subset \mathcal{A}$ eine $\mathcal{F}$-messbare, nicht-negative Zufallsgröße W mit $\mathbb{E}(X_n|\mathcal{F}) \nearrow W$ P-f.s. Der Satz von der monotonen Konvergenz und (II) liefern

$$\int_A X\, dP = \lim_{n\to\infty} \int_A X_n\, dP = \lim_{n\to\infty} \int_A \mathbb{E}(X_n|\mathcal{F})\, dP = \int_A W\, dP$$

für alle $A \in \mathcal{F}$. W erfüllt somit (I) und (II) und wir setzen $\mathbb{E}(X|\mathcal{F}) := W$. Diese Definition stimmt für quadratisch integrierbare X mit der vorher gegebenen Definition gemäß (c) in Lemma 13.2 überein. $\mathbb{E}(X|\mathcal{F})$ ist durch (I) und (II) P-f.s. eindeutig bestimmt (einfach)! Mit $A = \Omega$ in (II) ist

$$X \text{ integrierbar } \Leftrightarrow \mathbb{E}(X|\mathcal{F}) \text{ integrierbar.}$$

Damit folgt für quasi-integrierbares X vermöge $X = X^+ - X^-$:

$$\mathbb{E}(X|\mathcal{F}) := \mathbb{E}(X^+|\mathcal{F}) - \mathbb{E}(X^-|\mathcal{F}) \; .$$

13.3 Satz *Es seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $\mathcal{F}$ eine Unter-σ-Algebra von $\mathcal{A}$ und $X, Y, X_1, X_2, \ldots$ nicht-negative bzw. integrierbare Zufallsgrößen. Dann gilt:*

(a) $\mathbb{E}(\alpha X + \beta Y|\mathcal{F}) = \alpha\mathbb{E}(X|\mathcal{F}) + \beta\mathbb{E}(Y|\mathcal{F})$ *P-f.s. für alle* $\alpha, \beta \geq 0$ *bzw.* $\in \mathbb{R}$.

(b) $X \leq Y$ *P-f.s.* $\Rightarrow \mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F})$ *P-f.s.*

(c) $\mathbb{E}\big(\mathbb{E}(X|\mathcal{F})\big) = \mathbb{E}X$ *und* $\big|\mathbb{E}(X|\mathcal{F})\big| \leq \mathbb{E}\big(|X|\,\big|\mathcal{F}\big)$ *P-f.s.*

(d) $X_n \nearrow X$ *P-f.s.*, $X_n \geq 0$, $\Rightarrow \mathbb{E}(X_n|\mathcal{F}) \nearrow \mathbb{E}(X|\mathcal{F})$ *P-f.s.*

(e) $X_n \to X$ *und* $\sup_{n\geq 1} |X_n| \leq Y$ *P-f.s.*
$\Rightarrow \quad \mathbb{E}(X_n|\mathcal{F}) \to \mathbb{E}(X|\mathcal{F})$ *P-f.s.*

Beweis: Lemma 13.2 und Integrationstheorie □

Wir werden die folgenden Bezeichnungen verwenden:

Falls $X = 1_A$, $A \in \mathcal{A}$, so schreiben wir

$$P(A|\mathcal{F}) := \mathbb{E}(1_A|\mathcal{F}) \; ,$$

und nennen $P(A|\mathcal{F})$ die *bedingte Wahrscheinlichkeit* von A gegeben $\mathcal{F}$.

Ist $\mathcal{F} = \sigma(A) = \{\Omega, A, A^c, \emptyset\}$, so schreibt man auch $\mathbb{E}(X|A)$ für $\mathbb{E}(X|\mathcal{F})$ und analog für $\mathcal{F} = \sigma((A_i)_{i\in I})$ auch $\mathbb{E}(X|(A_i)_{i\in I})$ und $\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$, sowie $\mathbb{E}(X|(Y_i)_{i\in I})$ im Fall $Y = (Y_i)_{i\in I}$.

13.4 Beispiele

(a) Es sei $\mathcal{F} = \{\emptyset, \Omega\}$. Aus (I) folgt, dass $\mathbb{E}(X|\mathcal{F})$ konstant ist und mit (II) folgt $\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}X$. Ist $\mathcal{F} \supset \sigma(X)$, so folgt $\mathbb{E}(X|\mathcal{F}) = X$ P-f.s., denn X ist $\mathcal{F}$-messbar und es gilt (II).

(b) Ist $\mathcal{F} = \sigma(A) = \{\Omega, A, A^c, \emptyset\}$ mit $P(A) \in (0,1)$, so muss $\mathbb{E}(X|\mathcal{F})$ wegen der $\mathcal{F}$-Messbarkeit auf A und A^c jeweils konstant sein. Mit (II) folgt

$$\mathbb{E}(X|A) = \Big(P(A)^{-1} \int_A X\, dP\Big) 1_A + \Big(P(A^c)^{-1} \int_{A^c} X\, dP\Big) 1_{A^c} \; .$$

Hier ist P-f.s. überflüssig, da $\emptyset$ die einzige P-Nullmenge in $\mathcal{F}$ ist. Analog: Ist $\Omega = \dot{\bigcup}_{j=1}^{n} A_j$, $A_j \in \mathcal{A}$, $j = 1, \dots, n$, so gilt

$$\mathbb{E}(X|A_1, \dots, A_n) = \sum_{j=1}^{n} \Big(P(A_j)^{-1} \int_{A_j} X \, dP \Big) 1_{A_j} \quad P\text{-}f.s.$$

Hierbei setzen wir $\infty \cdot 0 := 0$, sofern $P(A_j) = 0$ für ein j. Sind alle $P(A_j)$ positiv, kann auf den Zusatz P-f.s. verzichtet werden.

13.5 Satz *Eine $\mathcal{F}$-messbare und integrierbare (nicht-negative) Zufallsgröße Y ist genau dann Version des bedingten Erwartungswertes $\mathbb{E}(X|\mathcal{F})$ einer integrierbaren (nicht-negativen) Zufallsgröße X, wenn*

$$\mathbb{E}(XZ) = \mathbb{E}(YZ)$$

für alle beschränkten (nicht-negativen) und $\mathcal{F}$-messbaren Zufallsgrößen Z gilt. Im Fall $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, $p \in [1, \infty]$, folgt letztere Behauptung sogar für alle $Z \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$ mit $\frac{1}{p} + \frac{1}{q} = 1$.

Beweis: Aus $\mathbb{E}(XZ) = \mathbb{E}(YZ)$ für alle Z wie angegeben folgt sofort (II), also auch $Y = \mathbb{E}(X|\mathcal{F})$ P-f.s., denn Y ist nach Voraussetzung $\mathcal{F}$-messbar. Ist umgekehrt $Y = \mathbb{E}(X|\mathcal{F})$ P-f.s., so folgt (II) und deshalb für nicht-negatives X auch $\mathbb{E}(XZ) = \mathbb{E}(YZ)$ für alle $\mathcal{F}$-messbaren, nicht-negativen Z, indem man Z durch eine monotone Folge aus $\mathcal{E}(\Omega, \mathcal{F})$ approximiert. Für integrierbares X erhält man die Aussage durch Zerlegung in Positiv- und Negativteil. Details bleiben als Übung. $\square$

13.6 Satz *Seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2$ Unter-σ-Algebren von $\mathcal{A}$ und X, Y Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$. Dann gilt:*

(a) *Sind X, Y nicht-negativ oder $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, $Y \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$ für $1 \leq p, q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, so gilt:*

$$Y \ \mathcal{F}\text{-messbar} \quad \Rightarrow \quad \mathbb{E}(XY|\mathcal{F}) = Y\mathbb{E}(X|\mathcal{F}) \ P\text{-f.s.}$$

(b) *Ist X quasi-integrierbar, so gilt:*

$$\mathcal{F}_1 \subset \mathcal{F}_2 \quad \Rightarrow \quad \mathbb{E}\Big(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1\Big) = \mathbb{E}\Big(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2\Big) = \mathbb{E}(X|\mathcal{F}_1) \ P\text{-f.s.}$$

(c) *Ist $\mathcal{F} = \sigma(\mathcal{F}_1 \cup \mathcal{F}_2)$, X quasi-integrierbar und die von X und $\mathcal{F}_1$ erzeugte σ-Algebra $\sigma(X, \mathcal{F}_1)$ unabhängig von $\mathcal{F}_2$, so gilt:*

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X|\mathcal{F}_1) \quad P\text{-f.s.}$$

(d) *Ist X quasi-integrierbar, sowie unabhängig von $\mathcal{F}$, so gilt:*

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}X \quad P\text{-f.s.}$$

Beweis:

(a) Seien X, Y nicht-negativ und Y $\mathcal{F}$-messbar. Dann gilt mit Satz 13.5 für alle nicht-negativen $\mathcal{F}$-messbaren Z

$$\int \mathbb{E}(XY|\mathcal{F})Z\,dP = \int XYZ\,dP = \int X(YZ)\,dP = \int \mathbb{E}(X|\mathcal{F})YZ\,dP\ ,$$

denn mit Z ist auch YZ nicht-negativ und $\mathcal{F}$-messbar. Da $\mathbb{E}(X|\mathcal{F})Y$ $\mathcal{F}$-messbar, folgt die Behauptung aus Satz 13.5. Für $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ und $Y \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$ ist XY integrierbar und $YZ \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$ für alle $\mathcal{F}$-messbaren und beschränkten Z. Der Beweis geht nun analog.

(b) Es gilt

$$\int_A \mathbb{E}\Big(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1\Big)\,dP = \int_A \mathbb{E}(X|\mathcal{F}_2)\,dP = \int_A X\,dP$$

für alle $A \in \mathcal{F}_1$ (und P-f.s. Eindeutigkeit des bedingten Erwartungswertes). Die zweite Gleichung folgt aus Beispiel 13.4 (a).

(c) Sei Y eine Version von $\mathbb{E}(X|\mathcal{F}_1)$. Diese ist $\mathcal{F}_1$- und somit auch $\mathcal{F}$-messbar. Wir zeigen, dass Y auch eine Version von $\mathbb{E}(X|\mathcal{F})$ ist. Dafür ist zu zeigen:

$$\int_A X\,dP = \int_A Y\,dP \text{ für alle } A \in \mathcal{F}\ . \tag{13.3}$$

Das System aller Mengen, welche diese Gleichheit erfüllen, ist ein Dynkin-System (Übung). Es genügt daher, (13.3) für alle A aus einem durchschnittstabilen Erzeuger von $\mathcal{F}$ zu zeigen. Wähle $\mathcal{E} := \{A_1 \cap A_2 \ :\ A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$. Wir bekommen dann

$$\begin{aligned}\int_{A_1\cap A_2} Y\,dP = \mathbb{E}(1_{A_1}1_{A_2}Y) &= P(A_2)\,\mathbb{E}(1_{A_1}Y)\\ &= P(A_2)\mathbb{E}(1_{A_1}X) = \int_{A_1\cap A_2} X\,dP,\end{aligned}$$

denn sowohl 1_{A_2} und $1_{A_1}Y$ als auch 1_{A_2} und $1_{A_1}X$ sind unabhängig. Wir verwenden Satz 10.21.

(d) Wir wählen in (c) $\mathcal{F}_1 = \{\Omega, \emptyset\}$. □

13.7 Satz (Jensensche Ungleichung) *Seien X eine integrierbare Zufallsgröße mit Werten in einem offenen Intervall $I \subset \mathbb{R}$ und $\varphi : I \to \mathbb{R}$ eine konvexe*

Funktion. Dann gilt $\mathbb{E}(X|\mathcal{F}) \in I$ P-f.s. für jede Unter-σ-Algebra $\mathcal{F} \subset \mathcal{A}$ und im Fall der Integrierbarkeit von $\varphi \circ X$ gilt

$$\varphi\big(\mathbb{E}(X|\mathcal{F})\big) \leq \mathbb{E}\big(\varphi(X)|\mathcal{F}\big) \quad P\text{-f.s.}$$

Beweis: Mit $P(X \in I) = 1$ folgt $P\big(\mathbb{E}(X|\mathcal{F}) \in I\big) = 1$, wobei wir die Monotonie, Satz 13.3 (b), verwenden: Aus $X(\omega) < \beta$ für alle $\omega \in \Omega$ folgt $\mathbb{E}(X|\mathcal{F}) \leq \beta$ P-f.s., also $\mathbb{E}(\beta - X|\mathcal{F}) \geq 0$ P-f.s. Nun ist $C := \{\mathbb{E}(\beta - X|\mathcal{F}) = 0\} \in \mathcal{F}$ und $\int_C (\beta - X)\, dP = 0$, somit $P(C) = 0$, also $\mathbb{E}(X|\mathcal{F}) < \beta$ P-f.s. Aus $\alpha < X(\omega)$ für alle $\omega \in \Omega$ folgt analog $\alpha < \mathbb{E}(X|\mathcal{F})$ P-f.s. In der Analysis lernt man, dass die konvexe Funktion φ in I rechts- und linksseitig differenzierbar ist, also auch stetig. Die rechtsseitige Ableitung $\varphi'_+(x)$ ist auf I isoton. Ferner gilt

$$\varphi(y) \geq \varphi(x) + \varphi'_+(x)(y - x)\ ,\ x, y \in I \tag{13.4}$$

Im Fall $x < y$ und $t \in (x, y)$ folgt dies aus

$$\frac{\varphi(x) - \varphi(t)}{x - t} \leq \frac{\varphi(x) - \varphi(y)}{x - y}$$

(was aus der Konvexität gefolgert werden kann). Für $x = y$ folgt Gleichheit in (13.4), also

$$\varphi(y) = \sup_{x \in I} \Big(\varphi(x) + \varphi'_+(x)(y - x)\Big), \quad y \in I\ .$$

Es gilt

$$\varphi(y) = \sup_{x \in I \cap \mathbb{Q}} \Big(\varphi(x) + \varphi'_+(x)(y - x)\Big), \quad y \in I\ , \tag{13.5}$$

denn φ'_+ ist isoton, also lokal beschränkt, und φ ist stetig. Also $\varphi \circ X \geq \varphi(x) + \varphi'_+(x)(X - x)$, und somit

$$\mathbb{E}(\varphi(X)|\mathcal{F}) \geq \varphi(x) + \varphi'_+(x)(\mathbb{E}(X|\mathcal{F}) - x)$$

P-f.s. für jedes $x \in I \cap \mathbb{Q}$. Wegen der Abzählbarkeit von $I \cap \mathbb{Q}$ gibt es ein $A \in \mathcal{F}$, $P(A) = 0$, mit

$$\mathbb{E}(\varphi(X)|\mathcal{F}) \geq \varphi(x) + \varphi'_+(x)(\mathbb{E}(X|\mathcal{F}) - x) \quad \text{auf } \Omega \setminus A$$

für jedes $x \in I \cap \mathbb{Q}$. Also folgt die Aussage auf $\Omega \setminus (A \cup B)$ mittels Anwendung von (13.5), wenn B ein Ereignis mit Wahrscheinlichkeit Null mit $\mathbb{E}(X|\mathcal{F})(\omega) \in I$ für alle $\omega \in \Omega \setminus B$ ist. $\square$

13.8 Korollar *Seien $p \geq 1$, $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ und $\mathcal{F} \subset \mathcal{A}$. Dann $|\mathbb{E}(X|\mathcal{F})|^p \leq \mathbb{E}(|X|^p|\mathcal{F})$ P-f.s., sowie $\mathbb{E}|\mathbb{E}(X|\mathcal{F})|^p \leq \mathbb{E}|X|^p$ (siehe Satz 13.3 (c)). Insbesondere ist $\mathbb{E}(X|\mathcal{F}) \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$.*

Beweis: $p = \infty$: mit X ist auch $\mathbb{E}(X|\mathcal{F})$ fast sicher beschränkt, siehe Satz 13.3 (c), sowie

$$X = \text{const.} = \alpha \quad \Rightarrow \quad \mathbb{E}(X|\mathcal{F}) = \alpha \; P\text{-f.s.}$$

$\square$

13.9 Satz *X, Y seien zwei Zufallsgrößen auf einem W-Raum $(\Omega, \mathcal{A}, P)$ und $\mathcal{F} \subset \mathcal{A}$ eine Unter-σ-Algebra. Dann gilt:*

(a) (Hölder-Ungleichung) *Aus $\mathbb{E}|X|^p < \infty$ und $\mathbb{E}|Y|^q < \infty$ für $1 \leq p, q \leq \infty$ mit $\frac{1}{p} + \frac{1}{q} = 1$ folgt*

$$\big|\mathbb{E}(XY|\mathcal{F})\big| \leq \mathbb{E}\big(|XY|\big|\mathcal{F}\big) \leq \Big(\mathbb{E}\big(|X|^p\big|\mathcal{F}\big)\Big)^{\frac{1}{p}} \Big(\mathbb{E}\big(|Y|^q\big|\mathcal{F}\big)\Big)^{\frac{1}{q}} < \infty \quad P\text{-f.s.}$$

(b) (Cauchy-Schwarz-Ungleichung) *Die Aussage aus (i) für $p = q = 2$.*

(c) (Minkowski-Ungleichung) *Aus $\mathbb{E}|X|^p < \infty$ und $\mathbb{E}|Y|^p < \infty$ für $p \geq 1$ folgt*

$$\Big(\mathbb{E}\big(|X+Y|^p\big|\mathcal{F}\big)\Big)^{\frac{1}{p}} \leq \Big(\mathbb{E}\big(|X|^p\big|\mathcal{F}\big)\Big)^{\frac{1}{p}} + \Big(\mathbb{E}\big(|Y|^p\big|\mathcal{F}\big)\Big)^{\frac{1}{p}} < \infty \quad P\text{-f.s.}$$

Beweis: Übung.

13.10 Bemerkungen

(a) Eine Beweisvariante zur Existenz bedingter Erwartungswerte verwendet den Satz von Radon-Nikodym, den wir bisher nicht betrachtet haben. Der Satz besagt, dass für zwei Maße μ und ν auf einem Messraum $(\Omega, \mathcal{A})$ (μ sei σ-endlich) ν eine Dichte bezüglich μ genau dann besitzt, wenn jede μ-Nullmenge auch eine ν-Nullmenge ist. X sei eine nicht-negative Zufallsgröße auf $(\Omega, \mathcal{A}, P)$, $\mathcal{F} \subset \mathcal{A}$ eine Unter-σ-Algebra. Es sei

$$Q_{X,\mathcal{F}}(A) := \int_A X \, dP \;, \quad A \in \mathcal{F},$$

ein Maß auf $(\Omega, \mathcal{F})$. Ist $P|_{\mathcal{F}}$ die Einschränkung von P auf $\mathcal{F}$, so ist jede $P|_{\mathcal{F}}$-Nullmenge auch eine $Q_{X,\mathcal{F}}$-Nullmenge. $P|_{\mathcal{F}}$ ist als W-Maß σ-endlich, also existiert nach dem Satz von Radon-Nikodym eine $\mathcal{F}$-messbare Zufallsgröße Y ($P|_{\mathcal{F}}$-f.s. eindeutig) mit

$$\int_A X \, dP = Q_{X,\mathcal{F}}(A) = \int_A Y \, dP|_{\mathcal{F}} = \int_A Y \, dP \;, \quad A \in \mathcal{F} \;.$$

Y besitzt also die Eigenschaften (I) und (II) einer Version des bedingten Erwartungswertes $\mathbb{E}(X|\mathcal{F})$.

(b) $P(A|\mathcal{F})$, wie nach Satz 13.3 eingeführt, ist eine $\mathcal{F}$-messbare, nicht-negative Funktion (f.s.) mit

$$\int_C P(A|\mathcal{F})\,dP = \int_C 1_A\,dP = P(A\cap C)\ , \quad C\in\mathcal{F}\ .$$

Wenn wir in der Situation von Beispiel 13.4 (b) sind, so folgt

$$P(A|A_1,\dots,A_n) = \sum_{j=1}^n \frac{P(A\cap A_j)}{P(A_j)} 1_{A_j} \quad P\text{-f.s.}$$

Es gilt $0 \le P(A|\mathcal{F}) \le 1$ f.s., $P(\emptyset|\mathcal{F}) = 0$ f.s., $P(\Omega|\mathcal{F}) = 1$ f.s. Aus $A_1 \subset A_2$, $A_1, A_2 \in \mathcal{A}$, folgt $P(A_1|\mathcal{F}) \le P(A_2|\mathcal{F})$ P-f.s. und für eine Folge $(A_n)_n$ paarweise fremder Ereignisse aus $\mathcal{A}$ gilt:

$$P\Big(\bigcup_{n\ge 1} A_n|\mathcal{F}\Big) = \sum_{n\ge 1} P(A_n|\mathcal{F}) \quad P\text{-f.s.} \tag{13.6}$$

(für jede Folge von Versionen der bedingten Wahrscheinlichkeiten $P(A_n|\mathcal{F})$ liefert die rechte Seite eine Version der bedingten Wahrscheinlichkeit auf der linken Seite!)

Es gilt nicht: Für fast alle $\omega\in\Omega$ ist

$$A \mapsto P(A|\mathcal{F})(\omega)$$

ein W-Maß auf $\mathcal{A}$! Denn in jeder der genannten Eigenschaften tritt eine von der betreffenden Gegebenheit, zum Beispiel in (13.6) von der Folge $(A_n)_n$ abhängige P-Nullmenge, als Ausnahmemenge auf. Die Vereinigung dieser im Allgemeinen überabzählbar vielen Mengen ist daher meist keine Nullmenge mehr!

Wir greifen das in (b) gestellte Problem auf und teilen inoffiziell mit:

13.11 Definition Es seien $(\Omega,\mathcal{A})$ und $(\Omega',\mathcal{A}')$ zwei Messräume. Eine Funktion $K:\Omega\times\mathcal{A}'\to\mathbb{R}$ heißt Markov*scher Kern* von $(\Omega,\mathcal{A})$ nach $(\Omega',\mathcal{A}')$, wenn

(a) $K(\cdot,A'):\Omega\to\mathbb{R}$ $\mathcal{A}$-messbar ist für jedes $A'\in\mathcal{A}'$

(b) $K(\omega,\cdot):\mathcal{A}'\to\mathbb{R}$ ein W-Maß auf $(\Omega',\mathcal{A}')$ für alle $\omega\in\Omega$ ist.

Es sei weiter $(\Omega,\mathcal{A},P)$ ein W-Raum und $\mathcal{F}$ eine Unter-σ-Algebra von $\mathcal{A}$. Ein Markovscher Kern von $(\Omega,\mathcal{F})$ nach $(\Omega,\mathcal{A})$ mit

$$K(\cdot,A) = P(A|\mathcal{F})\ P\text{-f.s.},\quad A\in\mathcal{A}\ ,$$

heißt *reguläre bedingte Wahrscheinlichkeit* bezüglich $\mathcal{F}$.

13.12 Satz *Ist Ω ein polnischer Raum, $\mathcal{A}$ die σ-Algebra der* BOREL*schen Mengen und P ein W-Maß auf $(\Omega, \mathcal{A})$. Dann existiert für jede σ-Algebra $\mathcal{F} \subset \mathcal{A}$ eine reguläre bedingte Wahrscheinlichkeit bezüglich $\mathcal{F}$. (ohne Beweis)*

Wir wenden uns noch $\mathbb{E}(X|Y) := \mathbb{E}\big(X\big|\sigma(Y)\big)$ zu. Sei X eine integrierbare reelle Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Weiter seien $Y : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ eine Zufallsvariable und $\mathbb{E}(X|Y)$ eine reelle integrierbare Version der bedingten Erwartung. Da $\mathbb{E}(X|Y)$ $\sigma(Y)$-messbar ist, existiert eine von der gewählten Version abhängige messbare reelle Funktion g auf $(\Omega', \mathcal{A}')$ mit

$$\mathbb{E}(X|Y) = g \circ Y\ . \tag{13.7}$$

Dies folgt aus dem

13.13 Lemma (Faktorisierungslemma) Sei $Y : \Omega \to \Omega'$ eine Abbildung der Menge Ω in den Messraum $(\Omega', \mathcal{A}')$ und $Z : \Omega \to \bar{\mathbb{R}}$ eine numerische Funktion auf Ω. Z ist genau dann $\sigma(Y)$-$\bar{\mathcal{B}}$-messbar, wenn es eine messbare numerische Funktion g auf $(\Omega', \mathcal{A}')$ gibt mit $Z = g \circ Y$.

Beweis: $Z = g \circ Y$ ist als Komposition einer $\sigma(Y)$-$\mathcal{A}'$-messbaren und einer $\mathcal{A}'$-$\bar{\mathcal{B}}$-messbaren Abbildung $\sigma(Y)$-$\bar{\mathcal{B}}$-messbar! Sei umgekehrt $Z = \sum_{j=1}^n \alpha_j 1_{A_j} \in \mathcal{E}\big(\Omega, \sigma(Y)\big)$, also $A_j \in \sigma(Y)$ und $\alpha_j \geq 0$, $j = 1, \dots, n$. Dann existieren $A'_j \in \mathcal{A}'$ mit $A_j = Y^{-1}(A'_j)$, und die Behauptung folgt mit $g = \sum_{j=1}^n \alpha_j 1_{A'_j}$. Jetzt folgt noch das Funktions-Erweiterungsargument (Übung). $\square$

13.14 Satz *Jede $\mathcal{A}'$-messbare reelle Funktion g, welche sich mit* (13.7) *aus einer reellen Version $\mathbb{E}(X|Y)$ ableitet, ist P^Y-integrierbar und es gilt*

$$\int_{A'} g\, dP^Y = \int_{\{Y \in A'\}} X\, dP\ , \quad A' \in \mathcal{A}' \tag{13.8}$$

Sie ist P^Y-fast sicher eindeutig bestimmt. Ist g eine $\mathcal{A}'$-messbare, reelle P^Y-integrierbare Funktion auf Ω', welche (13.8) *erfüllt, so ist $g \circ Y$ eine Version von $\mathbb{E}(X|Y)$.*

Beweis: Mit X ist $\mathbb{E}(X|Y)$ P-integrierbar und der Transformationssatz liefert für jedes $A' \in \mathcal{A}'$

$$\int_{\{Y \in A'\}} X\, dP = \int_{\{Y \in A'\}} \mathbb{E}(X|Y)\, dP = \int_{\{Y \in A'\}} g \circ Y\, dP = \int_{A'} g\, dP^Y\ ,$$

also ist g P^Y-integrierbar. Damit gilt (13.8). Ist h eine weitere $\mathcal{A}'$-messbare reelle Funktion mit $\mathbb{E}(X|Y) = h \circ Y$, so folgt $\int_{A'} g\, dP^Y = \int_{A'} h\, dP^Y$, $A' \in \mathcal{A}'$.

g und h sind P^Y-integrierbar und

$$\int_{A'} (g^+ + h^-)\, dP^Y = \int_{A'} (g^- + h^+)\, dP^Y \ , \quad A' \in \mathcal{A}' \ ,$$

also $g^+ + h^- = g^- + h^+$, also $g = h$ P^Y-f.s.

Gilt umgekehrt (13.8) für g, reell und P^Y-integrierbar, so ist $g \circ Y$ $\sigma(Y)$-messbar. Es liegt eine Version von $\mathbb{E}(X|Y)$ vor, denn

$$\int_{\{Y \in A'\}} g \circ Y \, dP = \int_{A'} g \, dP^Y = \int_{\{Y \in A'\}} X \, dP \ , \quad A' \in \mathcal{A}' \ ,$$

also $\int_C g \circ Y \, dP = \int_C X \, dP$ für alle $C \in \sigma(Y)$. □

Sei nun $\{y\} \in \mathcal{A}'$, $y \in \Omega'$, so liefert (13.8):

$$g(y) P(Y = y) = \int_{\{Y = y\}} X \, dP.$$

Wenn $P(Y = y) > 0$, folgt

$$g(y) = \frac{1}{P(Y = y)} \int_{\{Y = y\}} X \, dP =: \mathbb{E}_{\{Y=y\}}(X) =: \mathbb{E}(X|Y = y)$$

(siehe unsere anfängliche Diskussion).

$$\Rightarrow \mathbb{E}(X|Y)(\omega) = \mathbb{E}(X|Y = y) \quad \text{für alle } \omega \in \{Y = y\}.$$

Auf $\{Y = y\}$ mit $P(Y = y) > 0$ ist somit jede Version der bedingten Erwartung $\mathbb{E}(X|Y)$ konstant und gleich der bedingten Erwartung von X unter der Bedingung $\{Y = y\}$.

Im allgemeinen ist jedoch $P(Y = y)$ für viele $y \in \Omega'$ gleich Null (Lebesgue-stetig verteilte Zufallsgrößen sind eine Beispielklasse). Aber der Wert von g an einer Stelle $y \in \Omega'$ steht zur Verfügung.

13.15 Definition Für integrierbares, reelles X sei g eine der Bedingung (13.8) genügende $\mathcal{A}'$-messbare, P^Y-integrierbare, reelle Funktion. Dann heißt $g(y)$ für jedes $y \in \Omega'$ *bedingte Erwartung* von X unter der Bedingung, dass Y gleich y ist, in Zeichen

$$\mathbb{E}(X|Y = y) := g(y) \ .$$

$\mathbb{E}(X|Y = y)$ ist eine Zahl! Nach Satz 13.14 gewinnt man aus der Vergabe von $g(y) = \mathbb{E}(X|Y = y)$ eine Version der bedingten Erwartung $\mathbb{E}(X|Y)$ in der Form $\omega \mapsto \mathbb{E}\big(X \big| Y = Y(\omega)\big)$. Also $\mathbb{E}(X|Y)(\omega) = \mathbb{E}\big(X \big| Y = Y(\omega)\big)$ für fast alle $\omega \in \Omega$. Manchmal kann man $y \mapsto \mathbb{E}(X|Y = y)$ berechnen:

13.16 Satz *X, Y seien Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$, deren gemeinsame Verteilung $P^{(X,Y)}$ eine W-Dichte f besitzt: $P^{(X,Y)} = f\lambda^2$, $f : \mathbb{R}^2 \to \mathbb{R}_+$ eine $\mathcal{B}^2$-messbare Funktion. Ferner sei X integrierbar und*

$$f_0(y) := \int_{\mathbb{R}} f(x,y)\,dx > 0 \quad \text{für alle } y \in \mathbb{R}\ .$$

Dann ist

$$\mathbb{E}(X|Y=y) = \frac{1}{f_0(y)} \int x f(x,y)\,dx \quad \text{für } P^Y\text{-fast alle } y \in \mathbb{R}\ . \tag{13.9}$$

Es gilt weiter

$$\mathbb{E}(X|Y) = \frac{1}{f_0(Y)} \int x f(x,Y)\,dx \quad P\text{-f.s.}$$

Beweis: Es gilt für $A' \in \mathcal{B}$

$$\begin{aligned}\int_{\{Y \in A'\}} X\,dP &= \int X 1_{A'}(Y)\,dP = \int X 1_{\mathbb{R}\times A'}(X,Y)\,dP \\ &= \int x 1_{\mathbb{R}\times A'}(x,y)\,P^{(X,Y)}\big(d(x,y)\big) = \int_{\mathbb{R}\times A'} x f(x,y)\,dx\,dy\ .\end{aligned}$$

Insbesondere gilt für $A' = \mathbb{R}$: $\mathbb{E}(X) = \int_{\mathbb{R}\times\mathbb{R}} x f(x,y)\,dx\,dy$, also $(x,y) \mapsto x f(x,y)$ λ^2-integrierbar, also $y \mapsto \int x f(x,y)\,dx$ λ^1-fast überall definiert und λ^1-integrierbar (Fubini).

Da $\int f\,d\lambda^2 = 1$, folgt analog $f_0(y) < \infty$ λ^1-fast überall. Also existiert eine messbare Funktion $g : \mathbb{R} \to \mathbb{R}$:

$$g(y) := \begin{cases} \frac{1}{f_0(y)} \int x\, f(x,y)\,dx, & \text{falls } y \in \mathbb{R} \setminus N\ , \\ 0, & y \in N\ , \end{cases}$$

mit $\lambda^1(N) = 0$ und N so, dass $\int |x| f(x,y)\,dx < \infty$ und $f_0(y) < \infty$ für alle $y \in N^c$. N ist auch eine P^Y-Nullmenge, denn

$$P^Y(N) = P\big((X,Y) \in \mathbb{R} \times N\big) = \int_{\mathbb{R}\times N} f(x,y)\,dx\,dy = 0\ .$$

Sei g P^Y-integrierbar, so gilt für alle $A' \in \mathcal{B}^1$:

$$\begin{aligned}\int_{A'} g\,dP^Y &= \int_{\{Y\in A'\}} g \circ Y\,dP = \int 1_{\mathbb{R}\times A'}(X,Y) g \circ Y\,dP \\ &= \int_{\mathbb{R}\times A'} g(y) f(x,y)\,dx\,dy = \int_{A'} g(y) \Big(\int_{\mathbb{R}} f(x,y)\,dx\Big)\,dy \\ &= \int_{A'} g(y) f_0(y)\,dy = \int_{A'} \int_{\mathbb{R}} x f(x,y)\,dx\,dy = \int_{\mathbb{R}\times A'} x f(x,y)\,dx\,dy.\end{aligned}$$

Mit der anfänglichen Rechnung folgt

$$\int_{A'} g\,dP^Y = \int_{\{Y\in A'\}} X\,dP \ .$$

Für $A' = \mathbb{R}$ und $|g|$ anstelle von g folgt analog

$$\begin{aligned}\int |g|\,dP^Y &= \int |g| f_0\,d\lambda^1 = \int \Big| \int x f(x,y)\,dy \Big|\,dx \\ &\le \int\int |x f(x,y)|\,dx\,dy = \mathbb{E}(|X|) < \infty \ ,\end{aligned}$$

also die P^Y-Integrierbarkeit von g. Nun ist g also eine Funktion $y \mapsto \mathbb{E}(X|Y = y)$ wie in Definition 13.15 und Satz 13.14 liefert die Behauptungen. □

13.17 Beispiel (X,Y) besitze eine $N\Big(0, \frac{1}{1-\rho^2}\binom{\sigma^2\ \rho\sigma^2}{\rho\sigma^2\ \sigma^2}\Big)$-Verteilung, mit Dichte

$$f(x,y) = \frac{(1-\rho^2)^{1/2}}{2\pi\sigma^2} \exp\Big(-\frac{x^2 - 2\rho xy + y^2}{2\sigma^2}\Big)$$

für $(\sigma,\rho) \in (0,\infty) \times (-1,1)$. Dann ist

$$\begin{aligned} f_0(y) &= \int_{\mathbb{R}} f(x,y)\,dx \\ &= \frac{(1-\rho^2)^{1/2}}{2\pi\sigma^2} \exp\Big(-\frac{y^2(1-\rho^2)}{2\sigma^2}\Big) \int_{\mathbb{R}} \exp\Big(-\frac{(x-\rho y)^2}{2\sigma^2}\Big)\,dx \\ &= \frac{(1-\rho^2)^{1/2}}{2\pi\sigma^2} \exp\Big(-\frac{y^2(1-\rho^2)}{2\sigma^2}\Big)(2\pi\sigma^2)^{1/2} \\ &= \Big(\frac{1-\rho^2}{2\pi\sigma^2}\Big)^{1/2} \exp\Big(-\frac{y^2(1-\rho^2)}{2\sigma^2}\Big) \ . \end{aligned}$$

Also ist Y $N\Big(0, \frac{\sigma^2}{1-\rho^2}\Big)$-verteilt. Gemäß (13.9) folgt nun

$$\mathbb{E}(X|Y=y) = \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{\mathbb{R}} x \exp\Big(-\frac{(x-\rho y)^2}{2\sigma^2}\Big)\,dx = \rho\, y \ ,$$

also $\mathbb{E}(X|Y) = \rho\, Y$ P-f.s. □

Literaturverzeichnis

[1] Heinz Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, fifth edition, 2002. ISBN 3-11-017236-4.

[2] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. ISBN 0-471-00710-2. A Wiley-Interscience Publication.

[3] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

[4] Leo Breiman. *Probability*. Addison-Wesley Publishing Company, Reading, Mass., 1968.

[5] Kai Lai Chung. *A course in probability theory*. Academic Press Inc., San Diego, CA, third edition, 2001. ISBN 0-12-174151-6.

[6] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. ISBN 0-521-00754-2. Revised reprint of the 1989 original.

[7] Richard Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996. ISBN 0-534-24318-5.

[8] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[9] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[10] H.-O. Georgii. *Stochastik*. Walter de Gruyter, Berlin, 2002.

[11] Peter Gänssler and Winfried Stute. *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin, 1977. ISBN 3-540-08418-5.

[12] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

[13] K. R. Parthasarathy. *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3. Academic Press Inc., New York, 1967.

[14] S. R. S. Varadhan. *Probability theory*, volume 7 of *Courant Lecture Notes in Mathematics*. New York University Courant Institute of Mathematical Sciences, New York, 2001. ISBN 0-8218-2852-2

Index